

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:11:39 ; Search time 14.7 Seconds  
(without alignments)  
666.304 Million cell updates/sec

Title: US-09-631-863A-2

Perfect score: 401

Sequence: 1 MRNKRVLTKRRSGRGQD.....DNCPAKASKSPAGNSPAPL 401

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.2	13	4	US-09-058-562-32
2	9	2.2	15	4	US-08-602-999A-348
3	9	2.2	17	4	US-08-602-999A-413
4	9	2.2	18	4	US-08-602-999A-371
5	9	2.2	18	4	US-08-602-999A-409
6	9	2.2	19	4	US-08-602-999A-408
7	9	2.2	20	4	US-08-602-999A-357
8	9	2.2	20	4	US-08-602-999A-358
9	9	2.2	20	4	US-08-602-999A-359
10	9	2.2	20	4	US-08-602-999A-368
11	9	2.2	22	1	US-08-240-712-34
12	9	2.2	22	1	US-08-443-890-34
13	9	2.2	23	4	US-08-602-999A-376
14	9	2.2	32	3	US-08-545-196B-63
15	9	2.2	87	4	US-09-314-268-137
16	9	2.2	106	4	US-09-314-268-134
17	9	2.2	170	5	PCT-US95-16806A-6
18	9	2.2	210	1	US-08-234-783-2
19	9	2.2	210	1	US-08-456-907-2
20	9	2.2	210	5	PCT-US95-05523-2
21	9	2.2	223	1	US-08-167-035-43
22	9	2.2	223	1	US-08-208-887A-43
23	9	2.2	223	2	US-08-539-005-43
24	9	2.2	278	3	US-08-545-196B-19
25	9	2.2	281	2	US-08-810-453-2
26	9	2.2	281	3	US-08-815-190A-2
27	9	2.2	281	4	US-09-290-640-25

```

28 9 2.2 281 4 US-09-479-524-3 Sequence 3, Appli
29 9 2.2 281 4 US-08-339-214-8 Sequence 8, Appli
30 9 2.2 281 4 US-08-339-214-30 Sequence 30, Appli
31 9 2.2 281 5 PCT-US95-00362-2 Sequence 2, Appli
32 9 2.2 294 1 US-08-612-986-6 Sequence 6, Appli
33 9 2.2 294 1 US-08-361-806A-6 Sequence 9, Appli
34 9 2.2 294 3 US-08-545-196B-9 Sequence 4, Appli
35 9 2.2 294 4 US-09-028-327-4 Sequence 21, Appli
36 9 2.2 311 4 PCT-US95-16806A-21 Sequence 66, Appli
37 9 2.2 311 4 US-09-179-558-66 Sequence 1, Appli
38 9 2.2 377 1 US-07-863-169A-1 Sequence 1, Appli
39 9 2.2 377 2 US-08-429-964-1 Sequence 1, Appli
40 9 2.2 377 3 US-07-935-087-1 Sequence 1, Appli
41 9 2.2 377 5 PCT-US93-08062-1 Sequence 1, Appli
42 9 2.2 429 1 US-08-234-783-4 Sequence 4, Appli
43 9 2.2 429 1 US-08-456-907-4 Sequence 4, Appli
44 9 2.2 429 5 PCT-US95-05523-4 Sequence 4, Appli
45 9 2.2 445 2 US-08-900-148-2 Sequence 2, Appli

```

## ALIGNMENTS

```

RESULT 1
US-09-058-562-32
; Sequence 32, Application US/09058562A
; Patent No. 6184356
; GENERAL INFORMATION:
; APPLICANT: Anderson, David C.
; APPLICANT: Mathews, Antony James
; APPLICANT: Stetler, Gary L.
; TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC HEMOGLOBINS
; FILE REFERENCE: BXTB 2087
; CURRENT APPLICATION NUMBER: US/09/058,562A
; CURRENT FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: US 08/240,712
; PRIOR FILING DATE: 1994-05-09
; PRIOR APPLICATION NUMBER: PCT/US92/09752
; PRIOR FILING DATE: 1993-05-13
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)
; OTHER INFORMATION: Xaa=1-3 Gly
; NAME/KEY: VARIANT
; LOCATION: (12)
; OTHER INFORMATION: Xaa=1-5 Pro
; NAME/KEY: VARIANT
; LOCATION: (13)
; OTHER INFORMATION: Xaa=1-3 Gly
; OTHER INFORMATION: Description of Artificial Sequence: linker
; NAME/KEY: HELIX
; LOCATION: (2)-(11)
; OTHER INFORMATION: Proline helix
US-09-058-562-32

```

```

Query Match 2.2% Score 9; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 48 ppppppppp 56
    |
Db 2 ppppppppp 10

```

```

RESULT 2
US-08-602-999A-348

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:11:39 ; Search time 14.7 Seconds  
(without alignments)  
666.304 Million cell updates/sec

Title: US-09-631-863a-2

Perfect score: 401  
Sequence: 1 MNKRVLKTKRRSRGSGOD.....DNCPAKSKSPAGNSAPL 401

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 2442594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.2	13	4 US-09-058-562-32	Sequence 32, Appl
2	9	2.2	15	4 US-08-602-999A-348	Sequence 348, App
3	9	2.2	17	4 US-08-602-999A-413	Sequence 413, App
4	9	2.2	18	4 US-08-602-999A-371	Sequence 371, App
5	9	2.2	18	4 US-08-602-999A-409	Sequence 409, App
6	9	2.2	19	4 US-08-602-999A-408	Sequence 408, App
7	9	2.2	20	4 US-08-602-999A-357	Sequence 357, App
8	9	2.2	20	4 US-08-602-999A-358	Sequence 358, App
9	9	2.2	20	4 US-08-602-999A-359	Sequence 359, App
10	9	2.2	20	4 US-08-602-999A-368	Sequence 368, App
11	9	2.2	22	1 US-08-240-712-34	Sequence 34, Appl
12	9	2.2	22	1 US-08-443-890-34	Sequence 34, Appl
13	9	2.2	23	4 US-08-602-999A-376	Sequence 376, App
14	9	2.2	32	3 US-08-545-196B-63	Sequence 63, Appl
15	9	2.2	87	4 US-09-314-268-137	Sequence 137, App
16	9	2.2	106	4 US-09-314-268-134	Sequence 134, App
17	9	2.2	170	5 PCT-US95-16806A-6	Sequence 6, Appl
18	9	2.2	210	1 US-08-234-783-2	Sequence 2, Appl
19	9	2.2	210	1 US-08-456-907-2	Sequence 2, Appl
20	9	2.2	210	5 PCT-US95-05523-2	Sequence 2, Appl
21	9	2.2	223	1 US-08-167-035-43	Sequence 43, Appl
22	9	2.2	223	1 US-08-208-887A-43	Sequence 43, Appl
23	9	2.2	223	2 US-08-539-005-43	Sequence 43, Appl
24	9	2.2	278	3 US-08-545-196B-19	Sequence 19, Appl
25	9	2.2	281	3 US-08-810-453-2	Sequence 2, Appl
26	9	2.2	281	3 US-08-815-190A-2	Sequence 2, Appl
27	9	2.2	281	4 US-09-290-640-25	Sequence 25, Appl

28	9	2.2	281	4 US-09-479-524-3	Sequence 3, Appl
29	9	2.2	281	4 US-08-339-214-8	Sequence 8, Appl
30	9	2.2	281	4 US-08-339-214-30	Sequence 30, Appl
31	9	2.2	281	5 PCT-US95-00362-2	Sequence 2, Appl
32	9	2.2	294	1 US-08-612-986-6	Sequence 6, Appl
33	9	2.2	294	1 US-08-361-806A-6	Sequence 6, Appl
34	9	2.2	294	3 US-08-545-196B-9	Sequence 9, Appl
35	9	2.2	294	4 US-09-028-377-4	Sequence 4, Appl
36	9	2.2	294	5 PCT-US95-16806A-21	Sequence 21, Appl
37	9	2.2	311	4 US-09-179-558-66	Sequence 66, Appl
38	9	2.2	377	1 US-07-863-169A-1	Sequence 1, Appl
39	9	2.2	377	2 US-08-429-964-1	Sequence 1, Appl
40	9	2.2	377	3 US-07-935-087-1	Sequence 1, Appl
41	9	2.2	377	5 PCT-US93-08062-1	Sequence 1, Appl
42	9	2.2	429	1 US-08-234-783-4	Sequence 1, Appl
43	9	2.2	429	1 US-08-456-907-4	Sequence 4, Appl
44	9	2.2	429	5 PCT-US95-05523-4	Sequence 4, Appl
45	9	2.2	445	2 US-08-900-148-2	Sequence 2, Appl

## ALIGNMENTS

```

RESULT 1
US-09-058-562-32
: Sequence 32, Application US/09058562A
: Patent No. 6184356
: GENERAL INFORMATION:
: APPLICANT: Anderson, David C.
: APPLICANT: Mathews, Antony James
: APPLICANT: Stettler, Gary L.
: TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC HEMOGLOBINS
: FILE REFERENCE: EXTB 2087
: CURRENT APPLICATION NUMBER: US/09/058, 562A
: PRIOR FILING DATE: 1998-04-13
: PRIOR APPLICATION NUMBER: US 08/240,712
: PRIOR FILING DATE: 1994-05-09
: PRIOR APPLICATION NUMBER: PCT/US92/09752
: PRIOR FILING DATE: 1993-05-13
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 32
: LENGTH: 13
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: (1)
: OTHER INFORMATION: Xaa-1-3 Gly
: NAME/KEY: VARIANT
: LOCATION: (12)
: OTHER INFORMATION: Xaa-1-5 Pro
: NAME/KEY: VARIANT
: LOCATION: (13)
: OTHER INFORMATION: Xaa-1-3 Gly
: NAME/KEY: HELIX
: LOCATION: (2)..(11)
: OTHER INFORMATION: Proline helix
: OTHER INFORMATION: Proline helix
US-09-058-562-32

Query Match 2.28; Score 9; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 48 PPPPPPPP 56
Db 2 PPPPPPPP 10

RESULT 2
US-08-602-999A-348

```

Sequence 348, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OQUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
NUMBER OF SEQUENCES: 467  
TITLE OF INVENTION: ISOLATING AND USING SAME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 348:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-348

Query Match 2.28; Score 9; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 0.081; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0;

QY 48 PPPPPPPP 56  
Db 1 PPPPPPPP 9

RESULT 3  
US-08-602-999A-413  
Sequence 413, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OQUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
NUMBER OF SEQUENCES: 467  
TITLE OF INVENTION: ISOLATING AND USING SAME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas

CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 413:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 17 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-413

Query Match 2.28; Score 9; DB 4; Length 17;  
Best Local Similarity 100.0%; Pred. No. 0.091; Indels 0; Gaps 0;  
Matches 9; Conservative 0; Mismatches 0;

QY 48 PPPPPPPP 56  
Db 5 PPPPPPPP 13

RESULT 4  
US-08-602-999A-371  
Sequence 371, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OQUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLKES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
NUMBER OF SEQUENCES: 467  
TITLE OF INVENTION: ISOLATING AND USING SAME  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 371:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-371

Query Match 2.2%; Score 9; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.096;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 5  
US-08-602-999A-409  
Sequence 409, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 409:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-409

Query Match 2.2%; Score 9; DB 4; Length 18;  
Best Local Similarity 100.0%; Pred. No. 0.096;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 6  
US-08-602-999A-408  
Sequence 408, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 408:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-408

Query Match 2.2%; Score 9; DB 4; Length 19;  
Best Local Similarity 100.0%; Pred. No. 0.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 7  
US-08-602-999A-357  
Sequence 357, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILIAM, Lawrence A.

APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 357:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-357

Query Match 2.2%; Score 9; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 PPPPPPPP 56  
|||||  
Db 5 PPPPPPPP 13

RESULT 8  
US-08-602-999A-358  
Sequence 358, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: QUILIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 358:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-358

Query Match 2.2%; Score 9; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 PPPPPPPP 56  
|||||  
Db 5 PPPPPPPP 13

RESULT 9  
US-08-602-999A-359  
Sequence 359, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: QUILIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 359:  
SEQUENCE CHARACTERISTICS:



LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-359

Query Match 2.2%; Score 9; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 48 PPPPPPPP 56  
|||||  
Db 5 PPPPPPPP 13

RESULT 10  
US-08-602-999A-368  
Sequence 368, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OUILLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: MISTOCK, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 368:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-368

Query Match 2.2%; Score 9; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 0.11;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 48 PPPPPPPP 56  
|||||  
Db 4 PPPPPPPP 12

RESULT 11  
US-08-240-712-34  
Sequence 34, Application US/08240712  
Patent No. 5599607  
GENERAL INFORMATION:

APPLICANT: ANDERSON, DAVID C.  
APPLICANT: MATHEWS, ANTONY JAMES  
APPLICANT: STETLER, GARY L.  
TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
TITLE OF INVENTION: HEMOGLOBINS  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/240,712  
FILING DATE: 09-MAY-1994  
CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/09752  
FILING DATE: 13-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, IVER P  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: ANDERSON-6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633

INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:

OTHER INFORMATION: /note= one or both of Gly  
residues 2 and 3 can be absent; any or all  
of Pro residues 16, 17, 18 and 19 can be  
absent; one or both of Gly residues 22 and  
OTHER INFORMATION: 23 can be absent  
US-08-240-712-34

Query Match 2.2%; Score 9; DB 1; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 48 PPPPPPPP 56  
|||||  
Db 4 PPPPPPPP 12

RESULT 12  
US-08-443-890-34

Sequence 34, Application US/08443890  
Patent No. 5739011  
GENERAL INFORMATION:

APPLICANT: ANDERSON, DAVID C.  
APPLICANT: MATHEWS, ANTONY JAMES  
APPLICANT: STETLER, GARY L.  
TITLE OF INVENTION: PRODUCTION AND USE OF MULTIMERIC  
TITLE OF INVENTION: HEMOGLOBINS

NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/443,890  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/240,712  
FILING DATE: 09-MAY-1994  
APPLICATION NUMBER: PCT/US92/09752  
FILING DATE: 13-MAY-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, TYER P  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: ANDERSON-6  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
OTHER INFORMATION: /note= one or both of Gly  
OTHER INFORMATION: residues 2 and 3 can be absent; any or all  
OTHER INFORMATION: of Pro residues 16, 17, 18 and 19 can be  
OTHER INFORMATION: absent; one or both of Gly residues 22 and  
US-08-443-890-34  
Query Match 2.2%; Score 9; DB 1; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PPPPPPPP 56  
|||||  
DB 4 PPPPPPPP 12

RESULT 13  
US-08-602-999A-376  
Sequence 376, Application US/08602999A  
Patent No. 6184205  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: THORN, Judith M.  
APPLICANT: OULLIAM, Lawrence A.  
APPLICANT: DER, Channing J.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: RIDER, James E.  
TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
TITLE OF INVENTION: ISOLATING AND USING SAME  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas

CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/602,999A  
FILING DATE: 16-FEB-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-202  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 376:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-602-999A-376

Query Match 2.2%; Score 9; DB 4; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.12;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PPPPPPPP 56  
|||||  
DB 5 PPPPPPPP 13

RESULT 14  
US-08-545-196B-63  
Sequence 63, Application US/08545196B  
Patent No. 6080577  
GENERAL INFORMATION:  
APPLICANT: MELKI, JUDITH  
APPLICANT: MUMICH, ARNOLD  
TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE  
TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,196B  
FILING DATE: 19-OCT-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: FARACT, C. J.  
REGISTRATION NUMBER: 32,350  
REFERENCE/DOCKET NUMBER: 2121-110P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
INFORMATION FOR SEQ ID NO: 63:

SEQUENCE CHARACTERISTICS:  
; LENGTH: 32 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-545-196B-63

Query Match 2.2%; Score 9; DB 3; Length 32;  
Best Local Similarity 100.0%; Pred. No. 0.16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 8 PPPPPPPP 16

RESULT 15  
US-09-314-268-137  
; Sequence 137, Application US/09314268  
; Patent No. 6346377  
; GENERAL INFORMATION:  
; APPLICANT: DOOTBAR, John  
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA  
; FILE REFERENCE: 3789/80902  
; CURRENT APPLICATION NUMBER: US/09/314,268  
; EARLIER FILING DATE: 1999-03-19  
; EARLIER APPLICATION NUMBER: 09/314,268  
; NUMBER OF SEQ ID NOS: 179  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 137  
; LENGTH: 87  
; TYPE: PRT  
; ORGANISM: Human papillomavirus type 36  
US-09-314-268-137

Query Match 2.2%; Score 9; DB 4; Length 87;  
Best Local Similarity 100.0%; Pred. No. 0.41;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPP 56  
|||||  
DB 77 PPPPPPPP 85

Search completed: August 13, 2002, 15:13:57  
Job time: 138 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:07:09 ; Search time 31.61 Seconds  
(without alignments)  
1409.067 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 2139  
Sequence: 1 MRNKRVLTKRRSRGGGD.....DNCPRKSSPAGNSPAPL 401

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- A\_Geneseq\_032802.\*
- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
  - 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
  - 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
  - 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
  - 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
  - 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
  - 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
  - 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
  - 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
  - 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
  - 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
  - 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
  - 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
  - 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
  - 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
  - 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
  - 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
  - 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
  - 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
  - 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
  - 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
  - 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1802	84.2	342	21	Human ORFX ORF1912
2	245.5	11.5	1607	22	Novel human diagno
3	191	8.9	502	22	Novel human diagno
4	159	7.4	146	20	Human alpha helica
5	159	7.4	146	22	Human cell cycle a
6	156	7.3	520	22	Protein encoded by
7	155	7.2	126	20	Human mature alpha
8	149.5	7.0	110	20	Human zalphal epit
9	149.5	7.0	168	21	Human zsig83 matur
10	149.5	7.0	184	21	Human zsig83 prote
11	139.5	6.5	142	22	Novel human diagno

12	139	6.5	113	22	AA60475	Human cell cycle a
13	137	6.4	144	20	AAV5927	Human myometrium t
14	135.5	6.3	1230	19	AAW48895	Candida albicans C
15	132	6.2	557	19	AAV20666	Human neurofilamen
16	130	6.1	676	22	ABB59726	Drosophila melanog
17	129	6.0	872	22	AB90753	Human shear stress
18	126.5	5.9	567	22	ABB64078	Drosophila melanog
19	126.5	5.9	567	22	ABB65901	Drosophila melanog
20	126.5	5.9	567	22	ABB66007	C albicans apoptos
21	126	5.9	1072	22	AAV70871	Drosophila melanog
22	125	5.8	905	18	AAW31186	Human p160 polypep
23	125	5.8	1135	18	AAW31185	Human p160 polypep
24	123	5.8	874	22	ABB62601	Drosophila melanog
25	123	5.8	1321	22	ABG25416	Novel human diagno
26	122.5	5.7	1316	22	ABG22997	Novel human diagno
27	122.5	5.7	2348	22	ABG10929	Novel human diagno
28	121.5	5.7	486	22	AAU30801	Novel human secret
29	121.5	5.7	529	22	ABB12368	Human bone marrow
30	120.5	5.6	1157	22	ABG19976	Novel human diagno
31	120.5	5.6	1743	22	ABG10928	Novel human diagno
32	120	5.6	80	20	AAV26056	Human zalphal epit
33	120	5.6	954	22	AAU14615	Novel bone marrow
34	119.5	5.6	622	22	ABB62816	Drosophila melanog
35	119.5	5.6	2954	20	AAV01632	Amino acid sequenc
36	119	5.6	505	19	AAW46889	Human Neural-Wisko
37	119	5.6	505	22	AAM52316	Human N-WASP prote
38	119	5.6	505	22	ABG67338	Amino acid sequenc
39	119	5.6	582	22	ABG16505	Novel human diagno
40	118.5	5.5	2293	21	AAV69197	Amino acid sequenc
41	118	5.5	818	18	AAW13386	Human protein ubiq
42	118	5.5	2175	22	ABB65698	Drosophila melanog
43	117.5	5.5	501	19	AAW46890	Rat Neural-Wiskott
44	117.5	5.5	501	22	AAV52319	Rat N-WASP protein
45	117.5	5.5	501	22	ABG67355	Amino acid sequenc

ALIGNMENTS

RESULT 1  
AAB42148  
ID AAB42148 standard; Protein: 342 AA.  
AC AAB42148;  
XX  
DT 08-FEB-2001 (first entry)  
XX  
DE Human ORFX ORF1912 polypeptide sequence SEQ ID NO:3824.  
XX  
KW Human: Open reading frame; ORFX: detection; cytotropic; hepatotropic;  
KW vulnary; antipsoriatic; antiparkinsonian; nontropic; neuroprotective;  
KW anticonvulsant; osteopathic; antithrombotic; immunosuppressant; cardiant;  
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
KW hypotensive; dermatological; immunosuppressive; antitumor; antitubercu;  
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;  
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
KW cholesterol ester storage; systemic lupus erythematosus; infection;  
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KW thrombosis; contraceptive.

XX Homo sapiens.  
OS  
PN WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
PA  
XX  
XX Shimkets RA, Leach M;  
PI  
XX WPI; 2000-602362/57.  
DR N-PSDB; AAC76357.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 11; Page 2977-2978; 5507pp; English.  
XX  
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;  
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
XX Sequence 342 AA;

Query Match 84.2%; Score 1802; DB 21; Length 342;  
Best Local Similarity 100.0%; Pred. No. 1.8e-137;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 NNNSKHTGHSACVPNMTERRRDELSEENLNREKVMKQSENNNLQSOVKLTEENTT 119  
DB 1 nnnnskhtgksacvpnmterrerrdelseennlnrekvmkqseennnlqsgvqkleteentt 60  
QY 120 LREQVEPTDEDDDDIELGAAAAAAPPPIEECPEDLPKFDGNDMLAPFMAQCQIF 179  
DB 61 lreqveptdeddddielgaaaaaapppieecpedlpekfdgndmlapfmaqcf 120  
QY 180 MEKSTRFSDVRVCFVTSMTGGAARWASAKLERSHYLMHNPAPFMEMKHVFPDQR 239  
DB 121 mekstrfsvdrvcfvtsmttgaarwasaklershylmhnpapfmmemkhvfedpqr 180  
QY 240 REVAKRIRLRGMSGVIDYSNAFQMAODLWNEPALIDOVHEGLSDHQLSELSLEV 299  
DB 181 revakrirrkgmsgvidysnafqmagldwnepalidqyheglshdhlqeelslshlev 240  
QY 300 AKSLISALIGQCIHTERRLARAARKPRSPRALVLPHIAHQVDPTPEVGGARMRLTQ 359  
DB 241 akslialigqcihierrlaraarkprspralvlphiahqvdpptevggarmrltq 300  
QY 360 EEKERRRKLNLCLYCGTGGHYADNCPAKASKSPAGNSPAPL 401  
DB 301 eekerrrrklnlclycgtggghyadncpakaskspagnspapl 342

RESULT 2  
ABG15099

ID ABG15099 standard; Protein; 1607 AA.  
XX AC ABG15099;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #15090.  
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WC200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS79286.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID No 45458; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 1607 AA;

Query Match 11.5%; Score 245.5; DB 22; Length 1607;  
Best Local Similarity 25.1%; Pred. No. 7.9e-11;  
Matches 99; Conservative 54; Mismatches 149; Indels 93; Gaps 15;

QY 9 TKKRRS- --RGQDPLGPHRSEATAGRSPT-----PVTGLGDCPPPPPP-- 54  
DB 282 tteatsgsvrgeagpasgp---aqekkeppsgplqemeelpdlli-qdmeepsqprk 336  
QY 55 ----PPNN-----NNNNSKHTGHSACVPNMTERRRDELSEENLNREK-----VMKQ 99  
DB 337 eiedppndlllqleescngsh-----qarqdpisgasdrmkcasvnpngaree 384

QY 100 SEENNLOS-----QVQLTEENTT-----LREQVEPTPEDEDD 134  
Db 385 qeatldlkesgretpqeqnqtehtaemamvrssiislyfrmqdlkeq-----grvae 439  
QY 135 IELGAAAAAPPPIEECEPEDLPKFDGNDPMDLAFMAQCQIFMEKSTRDFSVDVRV 194  
Db 440 ilkginagqlp-----apkhfgsgdrrefhefivlcqitlqsgyrmfyndrlrv 488  
QY 195 CFVTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKHVFEDPQRRVAKRKIRRLRQGM 254  
Db 489 gyvinhisglalewakallgenspligdfafleamsevfeyrgalrvaeamftrqgg 548  
QY 255 GSVIDYNAFQMIQAQDLWDNEPALIDQYHEGLSDHIQELSHLEVAKSLSALIGQCIIH-I 313  
Db 549 rsateyidefgslvpilgwpdevlqahlcqglneeirhyl--frvpqdsrrqsdcapha 606  
QY 314 ERRLLA---RAAARKPRSPRALVPHIAHHQVD 345  
Db 607 nrreagreksyaqaagrppepdldrltcsrevd 641  
RESULT 3  
ABG09602  
ID ABG09602 standard; Protein; 502 AA.  
XX AC ABG09602;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #9593.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS73789.  
XX CC New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX CC Claim 20; SEQ ID NO 39961; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 502 AA;  
Query Match 8.9%; Score 191; DB 22; Length 502;  
Best Local Similarity 25.2%; Pred. No. 4.4e-07;  
Matches 55; Conservative 43; Mismatches 110; Indels 10; Gaps 3;  
QY 158 LPEKFDGNDPMDLAFMAQCQIFMEKSTRDFSVDVRVCFVTSMTGTGAARWASAKLERSH 217  
Db 21 lprlragspglmpveeemeideekgmkgflddsermaflvslhlgaerwflilqmevge 80  
QY 218 YLMHNYPAFMEMKHVFEDPQRRVAKRKIRRLRQGMGVSIDYSNAFOMIAQDLWDNEPA 277  
Db 81 plshenksfrrsggiydsleidilsavichpkqgksvryatdfillarhiswsdai 140  
QY 278 LIDQYHEGLSDHIQELSH--LEVAKLSLSALIGQCIIHERLARAAAAARKPRSPRALVL 335  
Db 141 lrtrfleglseavtkmgrriflkvagslkelidrslytecqlae----ekdspgnssqvl 196  
QY 336 PHIAHHQVDTEPVGGARMRLTQEEERRRKLNLCLY 373  
Db 197 ptaekrn----neeamgneissqgqteehqhvkrcyy 230  
RESULT 4  
AAAY26051  
ID AAAY26051 standard; Protein; 146 AA.  
XX AC AAAY26051;  
XX DT 28-SEP-1999 (first entry)  
XX DE Human alpha helical precursor protein-1, zalphal.  
XX KW Human alpha helical precursor protein-1; zalphal; helical cytokine;  
KW growth hormone; erythropoietin; leptin; chromosome Xq27.3; FMRI;  
KW Fragile-X syndrome; interleukin-10; connective tissue dysfunction;  
KW abnormal proliferative disorder; cancer; epidermal system; skin tone;  
KW hypothalamic pituitary gonadal axis dysfunction; cardiovascular system;  
KW cosmetic improvement; elasticity.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Peptide 1..20 /label= signal\_peptide  
FT Protein 21..146 /label= zalphal  
FT Region 23..37 /label= Helix\_A  
FT Region 53..67 /label= Helix\_B  
FT Region 82..96 /label= Helix\_C  
FT Region 118..132 /label= Helix\_D  
XX WO9929720-A2.  
XX PD 17-JUN-1999.  
XX PF 10-DEC-1998; 98WO-US26273.  
XX PR 10-DEC-1997; 97US-0987926.  
XX















[illegible]

Search completed: August 13, 2002, 15:11:11  
Job time: 242 sec

Query Match 6.28; Score 132; DB 19; Length 557;



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:07:34 ; Search time 14.65 Seconds  
(without alignments)  
668.578 Million cell updates/sec

Title: US-09-631-863A-2  
 Perfect score: 2139  
 Sequence: 1 MRNKRVLKTKKRSGRGGD.....DNCPAKASKSSPAGNSPAPL 401

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCUTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	159	7.4	146	4	US-09-209-525-2	Sequence 2, Appli
2	155	7.2	126	4	US-09-209-525-45	Sequence 45, Appl
3	149.5	7.0	110	4	US-09-209-525-52	Sequence 52, Appl
4	125	5.8	905	4	US-08-574-959A-9	Sequence 9, Appli
5	125	5.8	905	4	US-09-357-014-9	Sequence 9, Appli
6	125	5.8	1135	2	US-08-574-959A-7	Sequence 7, Appli
7	125	5.8	1135	4	US-09-357-014-7	Sequence 7, Appli
8	120	5.6	80	4	US-09-209-525-54	Sequence 54, Appl
9	118.5	5.5	2293	4	US-09-368-590-2	Sequence 6, Appli
10	118	5.5	834	3	US-08-539-205A-6	Sequence 6, Appli
11	115	5.4	1651	4	US-09-540-245A-18	Sequence 18, Appl
12	114	5.3	640	3	US-09-026-343-2	Sequence 2, Appli
13	112	5.2	1612	3	US-08-545-860D-48	Sequence 48, Appl
14	112	5.2	1612	5	PCR-US94-04496-48	Sequence 48, Appl
15	110.5	5.2	538	4	US-09-370-368-9	Sequence 9, Appli
16	109.5	5.1	1939	4	US-09-310-187A-1	Sequence 1, Appli
17	107	5.0	418	2	US-09-026-587-1	Sequence 1, Appli
18	107	5.0	418	2	US-09-227-420-1	Sequence 1, Appli
19	107	5.0	1829	4	US-09-157-420-1	Sequence 1, Appli
20	106.5	5.0	1099	4	US-09-442-100-2	Sequence 2, Appli
21	106	5.0	1872	1	US-08-188-582-14	Sequence 14, Appl
22	106	5.0	1872	1	US-08-646-715-14	Sequence 14, Appl
23	106	5.0	1886	4	US-08-938-105-3	Sequence 3, Appli
24	105.5	4.9	657	4	US-09-370-368-7	Sequence 7, Appli
25	104.5	4.9	393	2	US-09-026-587-3	Sequence 3, Appli
26	104.5	4.9	393	2	US-09-227-420-3	Sequence 3, Appli
27	104	4.9	2414	1	US-08-227-536-2	Sequence 2, Appli

```

:
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, Suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
:
```

COMPUTER READABLE FORM. disk  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/574,959A  
FILING DATE: 19-DEC-95  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: DFN-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEO ID NO: 9:

```

;
; 100 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
US-08-574-959A-9

Query Match          5.8%;   Score 125;   DB 2;   Length 905;
Best Local Similarity 20.3%;   Pred. No. 0.0055;
Matches 45;   Conservative 22;   Mismatches 75;   Indels 78;   Gaps 7;

QY      22  GLPHRSEATAGRPPTVTLGP-----DCPPPPPPPPPP--- 56
      || : || : || ||
Db      560  GLPLPPPPSGATPP-PIATGPPTASPVPAAKEEPEELPAAGPLPPPPPPPPVGP 618

QY      57  -----NNNNNNKHTGHKSACVPMWTRRRDE-- 84
      ||::: : || : ||
Db      619  VXLPPPLVPETPGGGGPPALEEDLTVINISDEEEEEESEEEEEEEEEEE 678

QY      85  -----LSEGINLRKVMKOSENNLOSOKLTENTTLRQVETPTDEDDDE-- 136

```

```

QY      137 -----LRGAAAAAAPPPIIEEC--PEDLPEKFGNDP 167
        |         |||||       :   |   :
Db      736 EDLEFGTAGVEEGAPPTLPALPPESPVKVQPEE 775

RESULT          5
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6281645
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;             and Jack L. Strominger
; TITLE OF INVENTION: p62 POLYPEPTIDES, RELATED POLYPEPTIDES
; AND USES THEREFOR

```

ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible

```

RESULT      5
US-09-357-014-9
; Sequence 9, Application US/09357014
; Patent No. 6291645
; GENERAL INFORMATION:
; APPLICANT: Jaekyoon Shin, Insil Joung, Ratna K. Vadlamudi
;              and Jack L. Strominger
; TITLE OF INVENTION: p52 POLYPEPTIDES, RELATED POLYPEPTIDES
;              AND USES THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```



```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-357-014-7

Query Match          5.8%; Score 125; DB 4; Length 1135;
Best Local Similarity 20.5%; Pred. No. 0.0075;
Matches 45; Conservative 22; Mismatches 75; Indels 78; Gaps 7;

Qy 22 GLPHRSATAGSPPTVTILGP-----DCPPPPPPPP----- 56
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 790 GLPLPPPPPGATPP-PIATGPTTAPPAKEPEELPAAPGLPPPPPPPPVPGP 848
Qy 57 -----NNNNNNSKHTGKHSACVPMNTERRDE-- 84
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 849 VXLPPPOLVPGTGGGGPPALEDLTVININSDSEEEEEESEEEEEEEEEE 908
Qy 85 -----LSEENINLRKVMKQSENNNNNQVOKLTTEENTTLREQVPTPEDEDDIE-- 136
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 909 EEEEDFEEDDEEYFESEEEEEEFEFEFE---EEGELEEEEEEDEEEEEELEEV 965
Qy 137 -----LRGAAAAAAPPPIIEEC--PEDLPEKFDGNPD 167
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 966 EDLEFGTAGGEVEGAPPPTLPALPPPPSPKVPQPEPE 1005

RESULT 8
US-09-209-525-54
; Sequence 54, Application US/09209525
; Patent No. 6303770
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Conklin, Darrell C.
; TITLE OF INVENTION: Mammalian Alpha Helical Protein-1
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/209,525
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 54
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-525-54

Query Match          5.6%; Score 120; DB 4; Length 80;
Best Local Similarity 32.9%; Pred. No. 0.00053;
Matches 25; Conservative 14; Mismatches 37; Indels 0; Gaps 0;

Qy 159 PEKFDGNPDMLAPMAQCIIFMEKSTRDFSDVRVRCVFTMTGGAARWASAKLERSHY 218
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 2 PETNGESSRLPEFIVQTASTVLYNENRNCNDAMKVAFLISLLAGEABEWWVPVPIEMDSP 61
Qy 219 LMHNYPAFMEMKHFV 234
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 62 ILGDYRAFLDEMCKCF 77

RESULT 9
US-09-368-590-2
; Sequence 2, Application US/09368590
; Patent No. 6187563
; GENERAL INFORMATION:
; APPLICANT: Solimena, Michele
; TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR
; FILE REFERENCE: 101918-200 (OCR-941)
; CURRENT APPLICATION NUMBER: US/09/368,590
; CURRENT FILING DATE: 1999-08-04
```

```
;
; EARLIER APPLICATION NUMBER: 60/095,657
; EARLIER FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2293
; TYPE: PRT
; ORGANISM: Human
US-09-368-590-2

Query Match          5.5%; Score 118.5; DB 4; Length 2293;
Best Local Similarity 23.3%; Pred. No. 0.079;
Matches 91; Conservative 41; Mismatches 145; Indels 113; Gaps 20;

Qy 29 EATAGRSPTTIV-----TLGPDPCPPPPPPPPPPNNNNNNNNNNKHTGKHSACVPMNTE 79
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1846 KAEQSKQPTTLLGRKFFGDPTLAAKAAPLLRP-----GGYERGLEP-LAR 1891
Qy 80 RRRDELSEENINLRKVMKQSENNNNNQVOKLTTEENTTLREQVEP-----TPDEDDDD 134
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1892 RASDTLSAEVT-RVGYYRQELKPERLQPRIDRLPE----IPGRVEPAALPAAPEDAET 1946
Qy 135 IELRGAAAAAAPPPIIEEC--PEDLPEKFDGNPDMLAPFMAQCIIFMEKSTRDFSDVRV 192
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1947 PATPAAAGQVRPRPERQESADRAEELPRR-----RRPERQESVDQS 1987
Qy 193 RVCFTVSMNTGRAARWASAKLERSHYLMHNYPAFMEMKHFVEDQORREVAKRKRRLRQ 252
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 1988 E-----EAAARRRRRPERQESAHEAHSLLTGR-----YEOMERRR--ERRRRLER 2031
Qy 253 GMGSVIDYSNAFQIAQDLWNEPALIDQYHEGLSDHIEQLSHLEVAKSLSALIGQCITH 312
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 2032 Q-----ESSEQEMPIRGLVKGKATLAD-----IVEQLOEKEAGGLPA--GPSLP 2075
Qy 313 IERRLAAAAARKPSPPRALVLPPIASHHQQVDPTPEPVGGARMRLTQBEKERRKLNLC 372
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 2076 QPREL-----PPGRLPNGLPELPTPR---PDRP--RARDRPKRRRRPRE----- 2117
Qy 373 YCGTGGHYADNCAPAKASKSPA--GNSPAP 400
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 2118 -GGEGG-----GSRRSKSAFAQGSAPAP 2140

RESULT 10
US-08-539-205A-6
; Sequence 6, Application US/08539205A
; Patent No. 6001619
; GENERAL INFORMATION:
; APPLICANT: Beach, David H.
; APPLICANT: Caligiuri, Maureen
; APPLICANT: Nefsky, Bradley
; TITLE OF INVENTION: Ubiquitin Ligases, and Uses Related Thereto
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/539,205A
; FILING DATE: 04-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: CSV-005.01
```

Query Match 5.3%; Score 114; DB 3; Length 640;  
Best Local Similarity 19.4%; Pred. No. 0.034;  
Matches 64; Conservative 55; Mismatches 125; Indels 86; Gaps 13;



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496  
FILING DATE:

## CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Deluca Esq., Mark

REGISTRATION NUMBER: 33,229

REFERENCE/DOCKET NUMBER: TJU-1242

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439

INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:

LENGTH: 1612 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

PCT-US94-04496-48

Query Match 5.2%; Score 112; DB 5; Length 1612;  
Best Local Similarity 19.7%; Pred. No. 0.19;  
Matches 60; Conservative 41; Mismatches 138; Indels 66; Gaps 9;

QY 31 TAGRSPTPTVTIGP---DCPPPPPPPPNNNNN-----NNSKHTGHSACV 74  
Db 1329 TPAATPATVAVSOPITDLPPTPPPPVHYAGDFDGMMDLPLPPPSANQIGLPSAQV 1388  
QY 75 PNMTERRRDELSEINNLREKVMKQSEENNLSQVQKLTETNTTLREQVEPTPEDEDD 134  
Db 1389 AAERKRREHQRWYE--KEKAPLEERERKRREQERKLGQMT---QSLNPAPFSPPLTA 1443  
QY 135 IELRGAAGAAAAPPP---TEEECPEDLPKFGDGNPMDLAPFMAQCQIFMEKSTRDFSVD 191  
Db 1444 QMKPEKPTLQRPQETVIRELPQQPRTIE-----RRDLQYITVSK 1486  
QY 192 VRVCFVTSMTTGAARWASAKLERSHYLMHNPAPFMMEMKHVFEDPORREVAKRIRRLR 251  
Db 1487 EELSSGDSLSPPDPKRWKRAKLEKQKQ--MHVMDLSKEIQELQSKPDGSAEESDRLRKLM 1545  
QY 252 QMGSGVIDYSNAPQMTAQDLDNNEPALIDQYHEGLSDHIQBELSHLEVAKSLSALIGQCI 311  
Db 1546 -----LEWQFQRLQESKQKDEDEDDDDVD-----TMLIMQRL 1581  
QY 312 HIERR 316  
Db 1582 EAERR 1586

## RESULT 15

US-09-370-368-9

Sequence 9, Application US/09370368

Patent No. 6258932

GENERAL INFORMATION:

APPLICANT: Anders Vahlne

TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY

FILE OF INVENTION: AND METHODS OF USE THEREOF

FILE REFERENCE: TRIPEP.003A

CURRENT APPLICATION NUMBER: US/09/370,368

CURRENT FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 9

LENGTH: 538

TYPE: PRT

ORGANISM: Moloney Murine Leukemia Virus

US-09-370-368-9

Query Match 5.2%; Score 110.5; DB 4; Length 538;  
Best Local Similarity 19.0%; Pred. No. 0.057;  
Matches 96; Conservative 51; Mismatches 184; Indels 175; Gaps 22;

QY 21 PGLHPHRS-----EATAGRSPTPTVTIGDPCPPPPPP-----PPNNNNN 61  
Db 71 PGPHGHPDQVPYIVTWEALAFDPPP---WVKPFVHPKPPPLLPSPAPSLPLEPPLS--- 123  
QY 62 NNSKHTGHSACVPMNTERRRDELSEINNLREKVMKQSEENNLSQVQKLTETNTTLR 121  
Db 124 -----TPPQSSLYPALTPSL-----GAKPKQVLSGGPL-----IDLTEDPPPYR 166  
QY 122 EQVEPTPEDEDDDI-ELRGAAGAAAAP-----PPIEECPEDLPKFDGPNPM 168  
Db 167 DP-RPPPSDRDGSGEATPAGEAPDFSPMASRLRGRREPPVADSTTSQAPPLRTGGNGQL 225  
QY 169 -LAPFMAQCQIFMEKSTRDFSVDVRV--CFVTSNMTGAAARW-----PPNNNNN 208  
Db 226 QYWPFSDDLWNKNNNPFSEDPGKLTALIESVLITHQPTWDDCQQLLGLLTGEEKQR 285  
QY 209 -----ASAKLER-----SHYLMH----- 221  
Db 286 VLLEARKAVRGDDGRPTQLPNEVDAAFPRLRPDWEYTTQAGRNHLVHYRQLLIAGLQNA 345  
QY 222 NYPAFMMEMKHVFEDPORREVAKRIRRLRQMGSVIDYS-----NAPQMTAQDLDNNEPA 277  
Db 346 RSTNLAKEVIGTQGP--NESPSAFLERKEAYRRYTPYDPDPGQGTNVSMFSIMOSAP 403  
QY 278 LIDQYHEGLSDHIQELSHLEVAKSLSALIGQCIHI-----ER 315  
Db 404 DIGRKLRL-----EDLRN---KTLGDLVREAEIRFNKRETPEREERIRREERKEER 454  
QY 316 RLARAAAARKPRPPRALVLPHTASHHQVDPTPEVGGARMRLTQEEKERRK---LNLCL 372  
Db 455 RRTEDQEKERDRRR-----HREMSRLLATVVSQORQDRQEGERRRRSLDCDOCT 505  
QY 373 YCOTGGHYADNCPAKASKSKSPAGNSP 398  
Db 506 YCEEQGHWAQDCPR--PRGPRGPRP 529

Search completed: August 13, 2002, 15:11:34

Job time: 240 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:08:59 ; Search time 20.9 Seconds  
(without alignments)  
1843.628 Million cell updates/sec

US-09-631-863A-2  
2139  
Title:  
Perfect score:  
Sequence: 1 MRNKRVLKTKKRRSGRGGD.....DNCPAKAKSSSPAGNSPAPL 401

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Database :      PIR_71:**
1: pir1:**
2: pir2:**
3: pir3:**
4: pir4:**
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	720.5	33.7	232	2	JF0163
2	240.5	11.2	853	2	S60178
3	190	8.9	457	2	T18347
4	159.5	7.5	639	2	S23569
5	148	6.9	1494	2	T13798
6	137	6.4	349	2	T18349
7	135	6.3	537	1	F0MVRV
8	134.5	6.3	555	2	T20349
9	134	6.3	1529	2	A9310
10	131.5	6.1	538	2	S70394
11	131.5	6.1	1230	2	T18256
12	131.5	6.1	1230	2	T18259
13	130	6.1	334	2	T01815
14	130	6.1	537	1	F0MVB
15	129	6.0	1110	2	T19673
16	129	6.0	1188	2	T46608
17	125	5.8	915	2	T26695
18	125	5.8	1651	2	T14160
19	123.5	5.8	1182	2	T30189
20	123.5	5.8	1585	2	T31611
21	122.5	5.7	536	1	F0MVE
22	122.5	5.7	538	2	S35474
23	122.5	5.7	601	2	S33377
24	122.5	5.7	1316	2	T00381
25	121.5	5.7	1187	2	T46637
26	121.5	5.7	1585	2	T18274
27	121	5.7	1612	2	T30805
28	120	5.6	745	2	D96829
29	120	5.6	747	2	S71478

30	120	5.6	992	2	T46337	hypothetical prote
31	120	5.6	1905	2	T18267	multidrug resistan
32	120	5.6	1937	2	T38055	myosin heavy chain
33	119.5	5.6	1870	2	S37671	MHC class III hist
34	119.5	5.6	1872	2	S36152	MHC class III hist
35	119.5	5.6	2954	2	T14156	kinesin-related pr
36	119	5.6	320	2	A52842	conserved hypothet
37	119	5.6	359	2	F57619	hypothetical prote
38	119	5.6	465	2	A02986	myosin alpha heavy
39	118.5	5.5	428	1	I36930	involucrin - white
40	118.5	5.5	7962	2	T38346	elastic titin - hu
41	118	5.5	2175	1	S03170	homeotic protein c
42	117	5.5	972	2	T49773	related to actin-i
43	116.5	5.4	249	2	A37280	C/EBP-related prot
44	116.5	5.4	708	2	I83136	NEDD-4 ORF - mouse
45	116.5	5.4	2142	2	B35098	MHC class III hist

## ALIGNMENTS

RESULT: 1  
JE0163  
myelin expression factor-3 - mouse  
N;Alternate names: MyEF-3  
C;Species: Mus musculus (house mouse)  
C;Date: 03-Jul-1998 #sequence\_revision 10-Jul-1998 #text\_change 07-May-1999  
C;Accession: JE0163  
R;Stępielewski, A.; Krynska, B.; Tretiakova, A.; Haas, S.; Khalili, K.; Amini, S.  
Biochem. Biophys. Res. Commun. 243, 295-301, 1998  
A;Title: MyEF-3, a developmentally controlled brain-derived nuclear protein which spe  
A;Reference number: JE0163; MUID:98139908

Query Match	33.7%;	Score	720.5;	DB	2;	Length	232;
Best Local Similarity	61.6%;	Pred. No.	1.6e-38;				
Matches	149;	Conservative	24;	Mismatches	52;	Indels	17;
Gaps	5;						
QY	154	CPEDLPEKFDGNDMLAPFMAOQCIPEWKEKSTRDVSVDVRVCFVTSMTWGRA--ARKWASA	211				
DB							
QY	2	CLEDLPEKFDGNDMLGPFMYOQLFMEKSTRDVSVDRIKRVCFVTSMLIGRAPLGYCAA	61				
DB							
QY	212	KL---ERSHYLMHNPAPFAMWKMKFVEDPQRREVAKRKIRRLRQGMGSVIDYSNAFQMI	267				
DB							
QY	62	KMYLPDAQHLHCLYDGAELVLR-----PSESVSVOTQDQTSAPGGPPVVDDYSNAFQMI	114				
DB							
QY	268	AQDLDWNEPALIDQYHEGLSDHIEQLSHLEVAKSLSALIGQCITHIERRLARAARAAAKPR	327				
DB							
QY	115	AQDLDWTEPALMDQOEGLNPDIPAELSRQEAEPKTLAALITACIHIERRLARADAAA-KPD	173				
DB							
QY	328	SPPRALVLPHIASHHQVDPTPEYGGARMRLTQBEKERRRKLNLCLYCGTGGHYADNCPAK	387				
DB							
QY	174	PSPRALVMP--FNSQTDPTPEYGGARMRLSKSEKERRRKMNLCLYCGNGGHFADTCPAK	230				
DB							
QY	388	AS	389				
DB							
QY	231	AS	232				
DB							

```

RESULT      2
S60178
gag polyprotein homolog - fungus (Fusarium oxysporum) retrotransposon skippy
C:species: Fusarium oxysporum
C:date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jul-2000
C:Accession: S60178

```



C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C;Accession: T13798  
R;Avedisov, S.  
submitted to the EMBL Data Library, February 1996  
A;Reference number: Z17761  
A;Accession: T13798  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1494 <AVE>  
A;Cross-references: EMBL:X95908; NID:e990667; PID:e223896; PIDN:CAA65152.1  
C;Genetics:  
A;Cross-references: FlyBase:FBgn0002698  
A;Mobile element: retrotransposon mdg3

Query Match	6.9%	Score 148;	DB 2;
Best Local Similarity	21.3%	Pred. No. 0.16;	Length 1494;
Matches 89;	Conservative	53;	Mismatches 155;
		Indels	120;
		Gaps	22;

[illegible]

RESULT 6  
T18349  
probable gag protein - rice blast fungus gypsy retroelement  
C:Species: Magnaporthe grisea (rice blast fungus)  
C:date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000  
C:Accession: T18349

K;DOBSON, K.F.  
submitted to the EMBL Data Library, September 1994  
A;Description: Sequence of the grh retroelement.  
A;Reference number: Z18883  
A;Accession: T18349  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-349 <DOB>  
A;Cross-references: EMBL:M77661; NID:g538065; PID:g538066; PIDN:AAA21441.1  
C;Genetics:  
A;Mobile element: gypsy retroelement  
C;Superfamily: rice blast fungus gypsy retroelement probable gag protein

Query Match 6.4%; Score 137; DB 2; Length 349;  
Best Local Similarity 23.3%; Pred. No. 0.14;  
Matches 80; Conservative 41; Mismatches 124; Indels 98; Gaps 16;  
QY 105 NLOSQVKLTENTTLREOVEPTPEDDDDIELRGAAAAAPP--PTTEECPEDLPEKF 162

[illegible]

```

Query Match      6.3%; Score 135; DB 1; Length 537;
Best Local Similarity 19.1%; pred. NO. 0.32;
Matches 97; Conservative 58; Mismatches 173; Indels 180; Gaps 20;

Qy      21 PGLPHRS-----RATAGRSP-----TPNVTUGDCP-----PPPPPPPPNNNN 61
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      71 PGPCHGDPQVPIVWCAIAYEPPSVKLSLSTAPILPSGFSQPPP----- 124

Qy      62 NNSKHTGHKSACVPMNTRRRDELSEINNLRKVMKQSENNNLQSOVKQLTEENTTLR 121
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      125 -----RSALYPALTP-----SIKPRSPQVLSDNGGLDILLTDEPPPYG 165

Qy      122 EQVEPTPEDDDDDTLRGAAAAAP-----PPTEERCPEDLPEKFDGNDM- 168
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      166 EQGSSPDGDDREAYTTSIIPAPSPMVSLRCKRPPPAANDSTTSRAFLRULGGGQLQ 225

Qy      169 LAPMAQCQIFMEKSTRDFSVDVRV-CFVTSMMTGAARWASAKLERSHYLMHNPAPM 227
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      226 YWPFSSDLYNWKNNPFSFEDPGKLTALTESVLTTHPTWDDCOQ-----L 272

```

QY	228	MEMKHVFEDPORREVAKRIRRLRQCGSVI--DYSNAEQMTAQDLDWNEP-----	276
DB	226	YWPFPSSDLYNWKNNPNPSEDPGKLTALLESVLTHTQPTWDDCCQ-----	271
DB	273	LGATLTAGEERQVLLKRAVRGNDGRPTQLPNEVNSAFPLRPDWDYTPPEGRNHLVY	332

QY 277 -----ALIDQYHEGL-----SDHQIE--- 292  
Db 333 ROLLLAGLQAGRSPTNLAKVGTQGNESPSAFLERLKEAYRYTPYPEDHGQTSV 392  
QY 293 -----ELSHLE--VAKSLSAIGQCIHI-----E 314  
Db 393 SMSFIQSAFDIGRLKLERLEDKSLTLDLVREAEIFNKRTPPEERFRFRTEENEE 452  
QY 315 RRLARAAARKPSPPR---ALVLPHIAHQVDTPVGGARMRLTOEKERRRKLNL 370  
Db 453 RRAADEQREKERRRRQREMSKLLATVVTGQRDRQ---GGERKR-PQLDKDQ----- 502  
QY 371 CLYCGTGCHVADNCPAKASKSSPAGNSP 398  
Db 503 CAYCKEGHAKDCPKK--PRGPRGPRP 528

RESULT 8  
T30349  
structural protein pp78-81 - Lymantria dispar nuclear polyhedrosis virus  
C:Species: Lymantria dispar nuclear polyhedrosis virus, LdMNPV  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T30349  
R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohd  
Virology 253, 17-34, 1999  
A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria d  
A:Reference number: 220836; MUID:99124785  
A:Accession: T30349  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-555 <KUZ>  
A:Cross-references: EMBL:AF081810; PIDN:AAC70187.1

Query Match 6.3%; Score 134.5; DB 2; Length 555;  
Best Local Similarity 20.7%; Pred. No. 0.35;  
Matches 75; Conservative 51; Mismatches 129; Indels 107; Gaps 16;

QY 21 PGLPHRSEATAGRSP-----TPTVTGLGDCPPPPPPPPNNNNNNKHTGKSCACV 74  
Db 214 PASEPARQESPIGSSAPEPIRQETPTGLFAPPPPPPPPPPP-----PPEPQOKSSAV 267  
QY 75 -----PNMTERRRDELSEINNLRKVMKQSENNNLSQVOK-----LTEENT 118  
Db 268 PPPPPPLPPPGAPDDPFGEIGQEV---RPKPAERAPTALFAEIRRGVQLKPATERAP 324  
QY 119 T-----LRE--QVEPTPEDED-----DDIELRGAAGAAAAPPP- 148  
Db 325 TYTPDALFAEIRQGVKLKPAERADEPPKPSRAPLLEIENRDKIKLKVAPRATPEPPA 384  
QY 149 -----PTE-----ECPEDLPKFDGNPDMPLAPFMAQCQIFMEKSTRDFSDV----- 190  
Db 385 SATNPLMQLLNKLRSEKMKSSAESDAN-----YTSSWSDAEDDSLRLDA 431  
QY 191 -RVRVCFTSMGTGR-----AARWASAKLERSHYLMHNYPAFMMEMKHVPEDPQREVAK 244  
Db 432 LRIKALLGPRLSERSEKRIAKRLAGAKLSAEXTLDLQARALEPDNPLSPPYQLTAP 491  
QY 245 RKTRRLRQMGSVID-----YSNAFOMI--AQDLDNWEPALIDQYHEGLSDHQIELSH 296  
Db 492 LYLHDLKLFSAVLDLFRNGAYETALEKLEALQVDLQAPSL-QRMHDDISTYVYKQKR 550  
QY 297 LE 298  
Db 551 LE 552

RESULT 9  
A59310  
unconventional myosin heavy chain - maize  
N:Alternate names: MYO1  
C:Species: Zea mays (maize)  
C>Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 08-Sep-2000

C:Accession: A59310  
R:Liu, L.; Pesacreta, T.C.  
submitted to Genbank, May 1999  
A:Reference number: A59310  
A:Accession: A59310  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1529 <LIU>  
A:Cross-references: GB:AF104924; NID:g4733890; PIDN:AAD17931.2; PID:g4733891  
C:Genetics:  
C:Gene: MYO1  
C:Superfamily: myosin MYO2; myosin motor domain homology  
F:65-719/Domain: myosin motor domain homology <MMO>

Query Match 6.3%; Score 134; DB 2; Length 1529;  
Best Local Similarity 22.3%; Pred. No. 1.2;  
Matches 73; Conservative 55; Mismatches 128; Indels 72; Gaps 14;

QY 78 TERREDELSEINNLRKVMKQSENNNLSQVOKLQTEENTTLREQ---VEP----- 126  
Db 1005 SELNEELIKKFFSAEKRIEQLQETVHRLEKATNMSENVKLRQQAIVASPTSKSLAAY 1064  
QY 127 -----TPEDEDDDIELRGAAGAAAAPPPPI-----EECPEDLPKFDGNPDML 169  
Db 1065 PKSPFQLKTPENG---ALNGEVKSSPDITPLPNPKLEAEERPKQSLNEKQEQENODLL 1121  
QY 170 APFMAQ-----CQIF-MEKSTRDFSDVRVCFVTSMMTGTAARWASAKLERS 216  
Db 1122 IKCVSQDLGFFSSGKPIAACLIYRCLLHWRSEFVERTGV-FDRIITQTGSAIESQDNNDKL 1180  
QY 217 HYLMHNPAPFMMEMKHVFE-----DPQREAVKRRRLRQGM-GSVIDYSNAF---Q 265  
Db 1181 AYWLSNSTLLLLQLORTLKTGAAGFTPQRRSSAAAFGRVFSGMRASPOGAFWMSR 1240  
QY 266 MIA-----QDLWNPEAL-----IDQYHEGLSDHQIELSHLEVAKSLSALIGQCIHIER 315  
Db 1241 LIGGLGLRQVEAKYPALLFKQLTAFLEKIYGMIRDNLK-----KEISPLGLGCIQAP- 1294  
QY 316 RLARAAAARKPSPPRALVLPFIASHHQ 343  
Db 1295 RTSRSLIKGSRQANALAQOTLIAHWQ 1322

RESULT 10  
S70394  
gag polyprotein - Friend murine leukemia virus (strain FB29)  
N:Contains: core protein p15; core shell protein p30; inner coat protein p12; nucleop  
C:Species: Friend murine leukemia virus  
A:Variety: strain FB29  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Jun-2000  
C:Accession: S70394  
R:Perryman, S.; Nishio, J.; Chesebro, B.  
Nucleic Acids Res. 19, 6950, 1991  
A:Title: Complete nucleotide sequence of Friend murine leukemia virus, strain FB29.  
A:Reference number: S70393; MUID:92107687  
A:Accession: S70394  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: genomic RNA  
A:Residues: 1-538 <PER>  
A:Cross-references: EMBL:211128; NID:g61547; PIDN:CAA74748.1; PID:g2654364  
A:Experimental source: strain FB29  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991  
C:Genetics:  
C:Gene: gag  
C:Superfamily: mammalian retrovirus gag polyprotein I  
C:Keywords: core protein; inner coat protein; nucleoprotein; polyprotein

Query Match 6.1%; Score 131.5; DB 2; Length 538;  
Best Local Similarity 18.8%; Pred. No. 0.53;  
Matches 95; Conservative 53; Mismatches 185; Indels 171; Gaps 17;







**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:10:39 ; Search time 13.43 Seconds  
(without alignments)  
1156.109 Million cell updates/sec

Title: US-09-631-863A-2

Perfect score: 2139

Sequence: 1 MRNKRVLTKTKRRSGRGQD.....DNCPAKASKSSPAGNSPAPL 401

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	159	7.4	146	1	LDOL_HUMAN
2	156	7.3	520	1	CET1_CANAL
3	135	6.3	537	1	GAG_MLVVD
4	131.5	6.1	538	1	GAG_MLVFF
5	131.5	6.1	1230	1	SP20_CANAL
6	130	6.1	537	1	GAG_MLVBM
7	129	6.0	872	1	S3B2_HUMAN
8	122.5	5.7	536	1	GAG_MLVDE
9	122.5	5.7	538	1	GAG_MLVFP
10	121.5	5.7	1585	1	P3K3_DICDI
11	120.5	5.6	1157	1	Y182_HUMAN
12	120.5	5.6	1816	1	AF6_HUMAN
13	120.5	5.6	1982	1	CHDM_DRONE
14	120	5.6	745	1	HGL2_ARATH
15	120	5.6	1905	1	TAGB_DICDI
16	120	5.6	1937	1	MYH8_HUMAN
17	119.5	5.6	1939	1	MYH1_HUMAN
18	119	5.6	465	1	MYH6_RABIT
19	119	5.6	505	1	WASL_HUMAN
20	118.5	5.5	428	1	INVO_CEBAL
21	118.5	5.5	2564	1	SPCQ_HUMAN
22	118	5.5	2175	1	HMCU_DRONE
23	117.5	5.5	501	1	WASL_RAT
24	116.5	5.4	281	1	CEBE_RAT
25	116.5	5.4	957	1	NED4_MOUSE
26	116.5	5.4	2142	1	BAT2_HUMAN
27	116	5.4	537	1	GAG_MLVAV
28	116	5.4	3680	1	DMD_CANFA
29	115.5	5.4	536	1	GAG_MLVCB
30	115	5.4	1227	1	B3A3_MOUSE
31	115	5.4	1332	1	SP71_YEAST
32	115	5.4	1935	1	MYH7_HUMAN
33	115	5.4	1935	1	MYH7_PIG

34	115	5.4	1939	1	MYH4_HUMAN	Q9V623	homo sapien
35	114.5	5.4	622	1	LAM0_DROME	P08928	drosophila
36	114.5	5.4	633	1	LA17_YEAST	O12446	saccharomyc
37	114.5	5.4	1941	1	MYH2_HUMAN	Q9UKX2	homo sapien
38	114	5.3	640	1	ELL2_HUMAN	O00472	homo sapien
39	113.5	5.3	281	1	CEBE_HUMAN	Q15744	homo sapien
40	113	5.3	1727	1	ALM1_SCHPO	Q9UTK5	schizosacch
41	113	5.3	1934	1	MYH7_MESAU	P13540	mesocricetu
42	112.5	5.3	2842	1	APC_RAT	P70478	rattus norv
43	112	5.2	502	1	WASP_HUMAN	P42768	homo sapien
44	112	5.2	1223	1	YWR1_CAEEL	Q10925	caenorhabdi
45	111.5	5.2	1232	1	B3A3_HUMAN	P48751	homo sapien

ALIGNMENTS

RESULT 1	LDOL_HUMAN	STANDARD;	PRT;	146 AA.
ID	LDOL_HUMAN			
AC	Q95751;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	LDOL1 protein (Leucine zipper protein down-regulated in cancer cells).			
GN	LDOL1			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=99330357; PubMed=10403563;			
RA	Nagasaki K., Manabe T., Hanzawa H., Maass N., Tsukada T.,			
RA	Yamaguchi K.;			
RT	"Identification of a novel gene, LDOL1, down-regulated in cancer cell			
RT	lines.";			
RL	Cancer Lett. 140:227-234(1999).			
CC	-!- FUNCTION: MAY HAVE AN IMPORTANT ROLE IN THE DEVELOPMENT AND/OR			
CC	PROGRESSION OF SOME CANCERS.			
CC	-!- SUBCELLULAR LOCATION: Nuclear.			
CC	-!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED WITH HIGH LEVELS IN			
CC	BRAIN ANT THYROID AND LOW EXPRESSION IN PLACENTA, LIVER AND			
CC	LEUCOCYTES, EXPRESSED AS WELL IN SIX OF THE SEVEN HUMAN BREAST			
CC	CANCER CELL LINES EXAMINED.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AB019527; BAA34364.1; -			
KW	Nuclear protein.			
FT	DOMAIN 132 143 ASP/GLU-RICH (HIGHLY ACIDIC).			
SQ	SEQUENCE 146 AA; 16968 MW; 3D02813B2DE52DBE CRC64;			

Query Match	7.4%;	Score 159;	DB 1;	Length 146;
Best Local Similarity	27.4%;	Pred. No. 0.0007;		
Matches	45;	Conservative 21;	Mismatches 68;	Indels 30; Gaps 2;
Qy	83	DELSEETNNLRKVMQSENNNIQSQVOKLFEENTTLREQVEPTPEDEDDIELGAAA	142	
		: : : :     : : : :     : : : :		
Db	3	DELVLLHALLMRHRLSIENSQMLEQLRLVCERASLLRQVRP-----	46	
Qy	143	AAAPPPPIEECPDLPEKFDGNDPMLAPFMAQCQIFMEKSTRDPSVDRVRCVFTSMWT	202	
		: : : :     : : : :     : : : :		
Db	47	-----PSCPVPFPPTFNGESSRLPEFIVQTASYMLVNLNENRRCNDAMKVAFLISLT	97	

```
Qy 203 GRAARWASAKLERSHYLMHNYPAFMMKMHVF-----EDPORRE 241
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 98 GEAEWVVPYIEMDSPILGDYRAFLDENKQCFGWDDDDDDDEE 141

RESULT 2
CET1_CANAL STANDARD; PRT; 520 AA.
AC O93803;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE mRNA capping enzyme beta subunit (Polynucleotide 5'-triphosphatase)
DE (EC 3.1.3.33) (mRNA 5'-triphosphatase) (TPase).
GN CET1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 1060;
RX MEDLINE=98427288; PubMed=9755857;
RA Yamada-Okabe H.; Mio T., Matsui M., Kashima Y., Arisawa M.,
RA Yamada-Okabe H.;
RT "Isolation and characterization of the Candida albicans gene for mRNA
RT 5'-triphosphatase: association of mRNA 5'-triphosphatase and mRNA 5'-
RT quanylyltransferase activities is essential for the function of mRNA
RT 5'-capping enzyme in vivo.";
RL FEBS Lett. 435:49-54(1998).
CC -1- FUNCTION: FIRST STEP OF M-RNA CAPPING. CONVERTS THE 5'-
CC -1- TRIPHOSPHATE END OF A NASCENT MRNA CHAIN INTO A DIPHOSPHATE END.
CC -1- CATALYTIC ACTIVITY: 5'-phosphopolynucleotide + H(2)O =
CC -1- polynucleotide + phosphate.
CC -1- COFACTOR: REQUIRES DIVALENT IONS (BY SIMILARITY).
CC -1- SUBUNIT: THE M-RNA CAPPING ENZYME IS COMPOSED OF TWO SEPARATE
CC CHAINS ALPHA AND BETA, RESPECTIVELY A MRNA GUANYLYLTRANSFERASE AND
CC AN RNA 5'-TRIPHOSPHATASE.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FUNGAL TPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB016242; BAA33965.1; -.
DR InterPro; IPR004206; mRNA_tripase.
DR Pfam; PF02940; mRNA_tripase; 1.
KW Hydrolase; mRNA processing; mRNA capping; Nuclear protein.
FT DOMAIN 92 95
FT POLY-SER
SQ SEQUENCE 520 AA; 58791 MW; 5C0690F547E13EBC CRC64;

Query Match 7.3%; Score 156; DB 1; Length 520;
Best Local Similarity 21.6%; Pred. No. 0.0048;
Matches 78; Conservative 59; Mismatches 118; Indels 106; Gaps 18;

Qy 9 TKKRRSGGGDDGLHPHRS-----EATAGRSPTT--PTVTLGPDGPPPP----- 50
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 24 TKIKSPATYHKPSVHERHSITSMLNDTPSDSTPTKKPEPTISPEFRKPSISLTSPSVA 83
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 51 --PPPPPPNNNNNSKHTGKSAVCPNWKERR-----RDELSEENLNRE 94
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 84 HKPPLPPSSSSVSSESSARSS--PAITKRNSTIANIIDAYEPAATKTEKKAELNSPK- 140
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy 95 KVMKQS-----EENNLIQSQVOKLTENTTLREQVEPTP--EDEDDDI-ELRGAAGAAA 145
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 141 --INQSTVPVKLEEHENDTNKVERKVDSPAEPKPKPEQPVPFDQDDDLTKIKLKQSKK 198
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
```





```
FT CHAIN 478 537 NUCLEOPROTEIN P10.
FT ZN_FING 501 518 CCHC-TYPE.
FT LIPID 2 MYRISTATE (BY SIMILARITY).
SQ SEQUENCE 537 AA; 60422 MW; ABD2E70299BFED64 CRC64;

Query Match 6.1%; Score 130; DB 1; Length 537;
Best Local Similarity 18.5%; Pred. No. 0.22;
Matches 93; Conservative 53; Mismatches 186; Indels 172; Gaps 17;

QY 21 PGLHPRS-----EATGRSPP-----TPVTVLGDCP-----PPPPPPNNNN 61
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 71 PGPHGPDQVPYIVTWEATAYEPWPVPFVSKLSPTAILSPGSTQPPP----- 124
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 62 NNSKHTGHSACVPNMTERRRRDELSEENLNRLKVKMKEEENNNLSQVOKLTTEHTLR 121
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 125 -----RSALYPAFT-----PSIKRPSKQVLSDDGGPL---IDLUTEPPPYG 165
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 122 EQVEPTDEDDIELGAAAAAAP-----PPPIEECEPELPEKFDGNPDW- 168
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 166 EQGPSSPDGDDREATSTSEIPAPSPMVSRLLRGRKDPDPAADSTTSRAFLRLGCGQLQ 225
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 169 LAPFMAQCIQFMEKTRDFSDVRVY-CFVTSMTTGRAARWASAKLERSHYLMHNPAPM 227
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 226 WPFSSDLYNWKNNPSPFSDPGKLTALIESVLTHOPTWDCCQ-----L 272
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 228 MEMKHVEDPQREVAKKIRLRQMGSVI--DYSNAFOMIAQDLDMNEP----- 276
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 273 LGTLTGEEKQVLEARKAVNGDGRPTQLPNEVNSAPFLRPDWDYTPBGRNHLVLY 332
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 277 -----ALIDQYHEGLSDH----- 289
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 333 RQLLAGLQAGRSPTNLAKYKITQGNPESPAFLERLKEAYRYTYPDPEDPQETNV 392
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 290 -----IQELSHLE--VAKLSALIGCQTHI-----E 314
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 393 SMSFIWQSAIPAIGRLERLEDLKSITGLDLVREAEKIFNKRETPEREERIRRETEE 452
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 315 RRLAAAAARKPRSPRALVLPPIASHHQVDPTPEVGGARMRLTOEEKERRRKLCLYC 374
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 453 RRRACDEQREKERRRRREMSKLTATVVTGQRQROGERRRQLDKDQ-----CAYC 506
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 375 GTGGHYADNCPAKASKSPAGNSP 398
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 507 KEKGHWKDCPKK--PRGRGPRP 528

RESULT 7
S3B2_HUMAN
ID S3B2_HUMAN STANDARD; PRT; 872 AA.
AC Q13435;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Splicing factor 3B subunit 2 (Spliceosome associated protein 145) (SAP
DE 145) (SF3b150) (Pre-mRNA splicing factor SF3b 145 kDa subunit).
GN SF3B2 OR SAP145
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N. A. AND SEQUENCE OF 151-159 AND 794-817.
RX MEDLINE=96154048; PubMed=8566756;
RA Gozani O., Feld R., Reed R.;
RT "Evidence that sequence-independent binding of highly conserved U2
RT snRNP proteins upstream of the branch site is required for assembly
RT of spliceosomal complex A."
RL Genes Dev. 10:233-243(1996).
RN [2]
RP CHARACTERIZATION OF THE SPLICEOSOME.
RX MEDLINE=20337962; PubMed=10882114;
RA Das R., Zhou Z., Reed R.;
```

```
RT "Functional association of U2 snRNP with the ATP-independent
RT spliceosomal complex E."
RL Mol. Cell 5:779-787(2000).
CC -!- FUNCTION: SUBUNIT OF THE SPLICING FACTOR SF3B REQUIRED FOR 'A'
CC COMPLEX ASSEMBLY FORMED BY THE STABLE BINDING OF U2 SNRNP TO THE
CC BRANCHPOINT SEQUENCE (BPS) IN PRE-MRNA. SEQUENCE INDEPENDENT
CC BINDING OF SF3A/SF3B COMPLEX UPSTREAM OF THE BRANCH SITE IS
CC ESSENTIAL. IT MAY ANCHOR U2 SNRNP TO THE PRE-MRNA. MAY ALSO BE
CC INVOLVED IN THE ASSEMBLY OF THE 'E' COMPLEX. BELONGS ALSO TO THE
CC MINOR U12-DEPENDENT SPLICEOSOME, WHICH IS INVOLVED IN THE SPLICING
CC OF RARE CLASS OF NUCLEAR PRE-MRNA INTRON.
CC -!- SUBUNIT: COMPONENT OF SPLICING FACTOR SF3B WHICH IS COMPOSED OF
CC FOUR SUBUNITS; SF3B4/SAP49, SF3B3/SAP130, SF3B2/SAP145,
CC SF3B1/SAP155. SF3B ASSOCIATES WITH THE SPLICING FACTOR SF3A AND A
CC 12S RNA UNIT TO FORM THE U2 SMALL NUCLEAR RIBONUCLEOPROTEINS
CC COMPLEX (U2 SNRNP). SF3B2 INTERACTS DIRECTLY WITH SF3B4.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO YEAST CUS1. SOME, TO C.ELEGANS ZK632.11.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U41371; AAA97461.1; -
CC MIM; 605591; -
CC InterPro; IPR003034; SAP.
CC Pfam; PF02037; SAP; 1.
CC SMART; SM00513; SAP; 1.
CC Spliceosome; mRNA processing; mRNA splicing; Nuclear protein.
FT DOMAIN 68 73 POLY-PRO.
FT DOMAIN 81 90 POLY-PRO.
FT DOMAIN 106 109 POLY-PRO.
FT DOMAIN 226 230 POLY-PRO.
FT DOMAIN 269 274 POLY-GLU.
FT DOMAIN 308 312 POLY-LYS.
FT DOMAIN 676 679 POLY-GLU.
FT DOMAIN 697 703 POLY-GLU.
SQ SEQUENCE 872 AA; 97656 MW; AED69FDD0DA5DE31 CRC64;

Query Match 6.0%; Score 129; DB 1; Length 872;
Best Local Similarity 20.7%; Pred. No. 0.45;
Matches 83; Conservative 64; Mismatches 150; Indels 104; Gaps 19;

QY 16 RGGQDPG-----LPHRSEATAGRSPTPTV-TILGPDCTPP--PPPPPPNNNNNNNSK 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 RPPQDMGQIGVPTLGP-RVAAPVGVGPTTVLPKGAAPVPRGPPPPGPDENREMDP 240
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 66 HTGHSACVPNMTER---RRDELSEENLNRLKVKMKEEENNNLSQVOKLTTEHTLR 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 SVGPK--IPQALEKITLQKESQEMNSQEEEMETDARSLSQASASETEEDTVSVSK 297
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 Q-----VEPTPEDEDDIELGAAAAAAPPPIEECEPELPE 160
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 298 KEKNRKRNRKKKKKQPRVGVSSSSGDRKXDSRGSPPAA---DVEIYVTEEP 354
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 161 KPDGN----PDLAPFMAQCQIFMEKSTRDFSVDRVRVCFVTSMGTGRAARWASAKLRS 216
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 355 IYEPNFIFFKRIFEAFKLTDDVKKKEKEPEKLD-----KLENS 393
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 217 HYLMMNYPAFMEMKHVPED-----PQREVAKKIRLRQMGSGVIDYSNAFOMI 267
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 394 --AAPKKGKFEHEKDDSDSDDEQEKKEAPKLSKKLRRMR-----FTVAELKQLV 446
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 268 AQDLDMNEPALIDQYHEGLSDHIEQLSHLEVAKSLSALIGCQIHLIERLARAAAAARKPR 327
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 447 AR-----PDVVEMHDVTAQD---PKLLVHLKATRN-SVPVPRHMCFKRKYLO--GKRGIE 495
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 328 SPPRALVLPPIASHHQVDPTPEVGGARMRLTOEEKERRRKL 368
```





```
QY 258 IDYSNAFOMIAQDLWNEPA-----LIDQYHEGLSDHIQEE-----LSHLEVAKSLS 304
Db 536 IOY-----NLNYPNTOKPTNIKLELV-----LEDELCKERLYDLOSLEINNGRP 580

QY 305 AL-----IGQCIIHERRLARAAARKPRS--PRALVLPH 337
Db 581 SIWKSHIDVLSFNKRLRELAKPQSNVPAARLTPY 618

RESULT 11
Y182_HUMAN
ID Y182_HUMAN STANDARD; PRT; 1157 AA.
AC Q14687;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0182 (Fragment).
GN KIAA0182.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (K1AA0161-K1AA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D80004; BA11499.1; -
KW Hypothetical protein.
FT NON_TER 1
FT DOMAIN 591 596 POLY-PRO.
FT DOMAIN 685 688 POLY-ARG.
FT DOMAIN 1042 1047 POLY-GLU.
FT SEQUENCE 1157 AA; 130323 MW; B0DADAESAFB5FB CRC64;
SQ -----

Query Match 5.6%; Score 120.5; DB 1; Length 1157;
Best Local Similarity 21.9%; Pred. No. 2.1;
Matches 95; Conservative 52; Mismatches 124; Indels 163; Gaps 25;

QY 23 LPHRSEATAGSPPTPTVT-----LGDPCPP-----PPPPP-----PNNNNNN 63
Db 354 LHGLRGHATEERGKPSQELTPTRAEKLDASLQAPKPVQHPHPTPHHTVPSLSNHG 413
QY 64 SKHTGHKACVPNTWTERDELSEINLRKVMQSENNNLQSOVKLFEENTTLREQ 123
Db 414 IFSLPSSGAATALLIORTNE-----EEKLARQRLRQEKEDR--QSQVSEFRQO--VLEQH 466
QY 124 VE---PTPEDDDIELRGAAAPPPPIEECPEDLPKFDGNDPDLAP-----FNAQ 175
Db 467 LDMGRPPVPAEAHREPE-----STTRPGNRHPEGRDPDPQHFGGPPPLISPKPOLHAAP 521
QY 176 QOIFMEKSTROFSVDRVRVCFVTSMTTGRAARWASAKLERSHYLMHNPAPF----- 226
Db 522 TALWNPVSLMD-----NTLETRA-----ESHSL-HSHPAAPPEPSRQAQV 560
QY 227 NMWEMKHVFDPQREVAKRKTRRLRQGMGSVIDYSNFAQMTAQDLWNNEPALIDQY--- 282
Db 561 PLVKVERVF-CPEKAEGRK-----REPAPLDKYQPP 592
```

```
QY 283 -----HEGLSDHIQEEELSH-----LEVAKLSALIGQCIIHERRLARAAAR--- 324
Db 593 PPPPPRGSG-----LEHQPFLPGPGPFLAELEKSTQTILGQ-----QRASLPQAATFGE 641
QY 325 -----KPRSP-----PRALVLPHIASSHQVDFT---EPVGGARMRLTQE---EKERRRL 368
Db 642 LSGPLKPGSPYRPPVPRA-----PDPAYIDFLOQRRRLVSKLDLEERRRREA 690
QY 369 NLCLYCGTGGHYAD 382
Db 691 Q-----EKGIYYID 698

RESULT 12
AF6_HUMAN
ID AF6_HUMAN STANDARD; PRT; 1816 AA.
AC P55196; O75087; O75088; O75089; Q9NU92;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE AF-6 protein.
GN MLLT4 OR AF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=94061833; PubMed=8242616;
RA Prasad R., Gu Y., Alder H., Nakamura T., Canaan O., Saito H.,
RA Huebner K., Gale R.P., Nowell P.C., Kuriyama K., Miyazaki Y.,
RA Croce C.M., Canaan E.;
RT "Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in
RT acute myeloid leukemias with the t(6;11) chromosome translocation.";
RL Cancer Res. 53:5624-5628(1993).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=98344142; PubMed=9679199;
RA Saito S., Matsushima M., Shirahama S., Minaguchi T., Kanamori Y.,
RA Minami M., Nakamura Y.;
RT "Complete genomic structure, DNA polymorphisms, and alternative
RT splicing of the human AF-6 gene.";
RL DNA Res. 5:115-120(1998).
RN [3]
RP SEQUENCE OF 337-1816 FROM N.A. (ISOFORM 2).
RA Williams S.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY ACT AS AN INTRACELLULAR SIGNALING COMPONENT
CC CONTROLLED BY RAS SIGNALING PATHWAYS.
CC -!- SUBUNIT: BINDS DIRECTLY TO ZO-1 AND OCCLUDIN.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1, 2 (SHOWN HERE) AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY A CHROMOSOMAL
CC TRANSLOCATION T(6;11)(Q27;Q23) THAT INVOLVES MLLT4 AND MLL/HRX.
CC THE RESULT IS A ROGUE ACTIVATOR PROTEIN.
CC -!- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011399; BAA32484.1; -
DR EMBL; AB011399; BAA32483.1; -
DR EMBL; AB011399; BAA32485.1; -
DR EMBL; U02478; AAC50059.1; -
```





CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF119716; AAD17276.1; -  
 DR EMBL; AE003515; AAF49099.1; ALT\_SEQ.  
 DR FlyBase: FBgn0013591; Mi-2.  
 DR InterPro: IPR000953; Chromo.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR002464; DEAH\_ATP\_helicase.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR InterPro: IPR001965; PHD.  
 DR InterPro: IPR000330; SNF2\_N.  
 DR InterPro: IPR001841; ZnF\_ring.  
 DR Pfam: PF00385; chromo; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR Pfam: PF00628; PHD; 2.  
 DR Pfam: PF00176; SNF2\_N; 1.  
 DR SMART; SM00298; CHROMO; 2.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELICC; 1.  
 DR SMART; SM00249; PHD; 2.  
 DR SMART; SM00184; RING; 2.  
 DR PROSITE; PS00598; CHROMO\_1; FALSE\_NEG.  
 DR PROSITE; PS00690; DEAH\_ATP\_HELICASE; 1.  
 DR PROSITE; PS00013; CHROMO\_2; 2.  
 DR DNA-binding; ATP-binding; Helicase; Nuclear protein; Repeat;  
 KW transcription regulation; Repressor; Zinc-finger.  
 FT ZN\_FING 380 421 PHD-TYPE.  
 FT ZN\_FING 440 481 PHD-TYPE.  
 FT DOMAIN 488 566 CHROMO 1.  
 FT DOMAIN 512 573 CHROMO 2.  
 FT NP\_BIND 755 762 ATP (POTENTIAL).  
 FT SITE 875 878 DEAH BOX.  
 FT DOMAIN 13 16 POLY-GLU.  
 FT DOMAIN 70 76 POLY-LYS.  
 FT DOMAIN 239 248 POLY-GLU.  
 FT DOMAIN 1279 1287 POLY-GLU.  
 FT DOMAIN 1672 1677 POLY-ASP.  
 FT MUTAGEN 737 737 G->D: IN ALLELE MI-2-5; LARVAL LETHAL.  
 FT CONFLICT 101 101 G -> A (IN REF. 1).  
 SQ SEQUENCE 1982 AA; 224199 MW; ED8E256D1AD0AC2F CRC64;  
 Query Match 5.6%; Score 120.5; DB 1; Length 1982;  
 Best Local Similarity 18.8%; Pred. No. 4;  
 Matches 80; Conservative 63; Mismatches 152; Indels 131; Gaps 16;  
 QY 27 RSEATAGRSP-----TPTVTLGDCPPPPPPPPNNNNNNKSHGKHSACVPM 77  
 Db 1529 KDVAALAEAPPNGVNDKATTSVTSATSAAPAPASEKGDKDKD----- 1577  
 QY 78 TERRDELSEETNNLRKVMKQSENN-----NLQSOVQKLT 114  
 Db 1578 SEKEDKTSAEKSEVKQE--QEAEDKPGDVQKQENPVVEAAGDKPSDAEVKTEVAKTE 1635  
 QY 115 ENTTLRQVEPTPDEDDDELGRGAAAAAPPPIEECECPDELPEKFDGNDMLAPFMA 174  
 Db 1636 KPEETKDPKEVKEEPEEKEKEKVDKRPPTPTVTTIDDDDDVMIVKEDGE----- 1686  
 QY 175 QCOIFMEKSTRDFSVDYRVCFVTSMTTCRAARWASAKLE--RSHVLMHNYPAFWMEMKH 232  
 Db 1687 -----LEKPSASSPKDKAVAAANTSATGATGAGDESLVUKRFMFNIADGGTELHT 1741  
 QY 233 VFEDPQRREVARKRTI-----LRQGMGSVIDYSN--AFQMIADLDWNEPA 277  
 Db 1742 LWLNKEKAAPGREGVEIWHRRHRYWLLAGIVTHGYGRQDIQNDIRFAII-----NEPF 1795  
 QY 278 LIDQVHEGLSDHIQELSHLEVAKLSA-----LIQCQIHIERRLRAAAARKPRSPRAL 333

Db 1796 KMD-VYKG-----NFEIKNFKLARRFKLLEQALVIEEQLRRAAYLNLAQ----- 1839  
 QY 334 VLPFIASHHQVDPTEPVGGARMRLTQEEKERRKLNLCLYCGTGGHYADNCAPAKSKSSP 393  
 Db 1840 -----DPSHPAMSLNARFAVE-----CLAESH-----QHLKSLE 1870  
 QY 394 AGNSPA 399  
 Db 1871 AGNKPA 1876  
 RESULT 14  
 ID HGL2\_ARATH STANDARD; PRT; 745 AA.  
 AC P46607; Q39018;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein GLABRA2 (Homeobox-leucine zipper protein ATHB-10)  
 DE (HD-ZIP protein ATHB-10).  
 GN GL2 OR AT1G79840 OR F19K16.20.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. WASSILEWSKIIJA; TISSUE=Seedling;  
 RX MEDLINE=95011550; PubMed=7926739;  
 RA Rerle W.G., Feldmann K.A., Marks M.D.;  
 RT "The GLABRA2 gene encodes a homeo domain protein required for normal  
 RL trichome development in Arabidopsis.";  
 RL Genes Dev. 8:1388-1399(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Marks M.D.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=96407838; PubMed=8811855;  
 RA di Cristina M., Sessa G., Dolan L., Linstead P., Baima S., Ruberti I.,  
 RA Morelli G.;  
 RT "The Arabidopsis Athb-10 (GLABRA2) is an HD-zip protein required for  
 RL regulation of root hair development.";  
 RL Plant J. 10:393-402(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.-J., Koo H.-U., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,  
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaya V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RL thaliana.";  
 RL Nature 408:816-820(2000).  
 CC -!- FUNCTION: REQUIRED FOR CORRECT MORPHOLOGICAL DEVELOPMENT AND  
 CC MATURATION OF TRICHOMES AS WELL AS FOR NORMAL DEVELOPMENT OF SEED



FT DOMAIN 1782 1785 POLY-SER.  
FT DOMAIN 1807 1812 POLY-PRO.  
FT DOMAIN 1813 1860 POLY-GLN.  
FT DOMAIN 1872 1878 POLY-PRO.  
FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 747 747 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 823 823 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1172 1172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1522 1522 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1658 1658 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1905 AA; 212518 MW; B8E223FA8B9AE13C CRC64;

Query Match 5.6%; Score 120; DB 1; Length 1905;  
Best Local Similarity 19.9%; Pred. No. 4.1;  
Matches 30; Conservative 25; Mismatches 52; Indels 44; Gaps 4;  
QY 26 HRSEATAGRPPTPTVTGLGDCPPP-----ppppppnnnnnnnSKHTGHSACVP 75  
Db 1777 NNNEPSSSTPPNDQPT-----PPQEQEQEKNDQPPPPPQEQ----- 1815  
QY 76 NMTERRRDELSEINNLREKVMKQSENNLQSQVKLTENTTLREQVEPTPEDEDDDI 135  
Db 1816 ---OQEQEQQQQQEQQQQQQQQQQQQQQQQQQQQQQQQQQQNDQPPNDYDQ-- 1870  
QY 136 ELRGAAAAAPPPPIEECPDLPEKFDGNP 166  
Db 1871 -----VPPPPPLPSESPPPTGNNDGQP 1893

Search completed: August 13, 2002, 15:13:00  
Job time: 141 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:09:24 ; Search time 31.91 Seconds  
(without alignments)  
2173.957 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 2139  
Sequence: 1 MRNKRVLTKTKRRSGRGQD.....DNCPSAKSKSSPAGNSPAPL 401

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	1802	84.2	342	4	Q9UPV1	Q9UPV1	homo sapien
2	1703	79.6	325	4	Q96A68	Q96A68	homo sapien
3	943	44.1	231	11	Q9EQ11	Q9EQ11	mus musculus
4	373	17.4	704	13	Q98SV9	Q98SV9	fugu rubrip
5	364.5	17.0	371	13	Q93283	Q93283	fugu rubrip
6	279	13.0	1252	3	Q96V43	Q96V43	aspergillus
7	240.5	11.2	853	3	Q00833	Q00833	fusarium ox
8	229.5	10.7	837	3	Q9HFY8	Q9HFY8	colletotric
9	223.5	10.4	400	6	Q95KK1	Q95KK1	macaca fasc
10	192.5	9.0	1418	4	Q9P2M8	Q9P2M8	homo sapien
11	189	8.8	1810	10	Q94115	Q94115	oryza sativ
12	185	8.6	1500	10	Q94DA0	Q94DA0	oryza sativ
13	185	8.6	1862	10	Q94104	Q94104	oryza sativ
14	184.5	8.6	1781	10	Q9AYC0	Q9AYC0	oryza sativ
15	182.5	8.5	1473	10	Q94H22	Q94H22	oryza sativ
16	180.5	8.4	959	10	Q94108	Q94108	oryza sativ

17	179.5	8.4	1714	10	Q94HT9	Q94HT9	oryza sativ
18	179.5	8.4	1777	10	Q94HP9	Q94HP9	oryza sativ
19	179	8.4	2017	10	Q9AYB6	Q9AYB6	oryza sativ
20	178.5	8.3	1491	10	Q9XE26	Q9XE26	oryza sativ
21	177.5	8.1	1773	10	Q94HV6	Q94HV6	oryza sativ
22	170.5	8.0	1506	10	Q94DR3	Q94DR3	oryza sativ
23	167	7.8	1524	10	Q9LWJ0	Q9LWJ0	oryza sativ
24	166.5	7.8	1475	10	Q94H45	Q94H45	oryza sativ
25	164.5	7.7	1571	10	Q94HN3	Q94HN3	oryza sativ
26	163	7.6	1521	10	Q9XEM5	Q9XEM5	oryza sativ
27	163	7.6	2157	10	Q9AYB5	Q9AYB5	oryza sativ
28	162.5	7.6	406	3	Q9UVD9	Q9UVD9	alternaria
29	161.5	7.6	2162	10	Q9AYC2	Q9AYC2	oryza sativ
30	161	7.5	1220	10	Q9XHY2	Q9XHY2	oryza sativ
31	161	7.5	1521	10	Q94DL1	Q94DL1	oryza sativ
32	159.5	7.5	639	3	Q9UVC2	Q9UVC2	cladosporiu
33	159.5	7.5	963	10	Q9LWX0	Q9LWX0	oryza sativ
34	157	7.3	1524	10	Q93VN0	Q93VN0	oryza sativ
35	156	7.3	1461	10	Q94H43	Q94H43	oryza sativ
36	151	7.1	372	3	Q9Y8H2	Q9Y8H2	tricholoma
37	148	6.9	1494	5	Q94885	Q94885	drosophila
38	146	6.8	1524	10	Q943A1	Q943A1	oryza sativ
39	142.5	6.7	303	4	Q9H7A0	Q9H7A0	homo sapien
40	139	6.5	1587	10	Q9SLQ0	Q9SLQ0	oryza sativ
41	138.5	6.5	2075	13	Q90WA4	Q90WA4	fugu rubrip
42	138	6.5	675	11	Q91YM9	Q91YM9	mus musculus
43	138	6.5	1123	11	Q9DBD5	Q9DBD5	mus musculus
44	137	6.4	498	10	Q94BM0	Q94BM0	hordeum vul
45	135	6.3	100	11	Q9DC23	Q9DC23	mus musculus

ALIGNMENTS

RESULT 1  
Q9UPV1  
ID Q9UPV1 PRELIMINARY; PRT; 342 AA.  
AC Q9UPV1;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE KIAA1051 PROTEIN (FRAGMENT).  
GN KIAA1051.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=93997452; PubMed=10470851;  
RA Kikuno K., Nagase T., Ishikawa K., Hirotsawa M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";  
RL DNA Res. 6:197-205(1999).  
DR EMBL; AB028974; BAA83003.1; -;  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00098; zf-CCHC; 1.  
DR SMART; SM00343; Znf\_C2HC; 1.  
KW Zinc-finger.  
FT NON\_TER  
SQ SEQUENCE 342 AA; 38769 MW; DC14C265B92541EB CRC64;

Query Match 84.2%; Score 1802; DB 4; Length 342;  
Best Local Similarity 100.0%; Pred. No. 7.8e-131;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 60 NNNNSKHTGKHSACVPNMTERRRDELSEETNNLRKVMKQSENNNLQSOVKLTTEENTT 119  
Db 1 NNNNSKHTGKHSACVPNMTERRRDELSEETNNLRKVMKQSENNNLQSOVKLTTEENTT 60

Qy	120	LREQVEPTPEDEDDIELRGAAAAAAPPPIIEECPEDLPEKFDGNPMDLAPFMAQCQIF	179
Db	61	LREQVEPTPEDEDDIELRGAAAAAAPPPIIEECPEDLPEKFDGNPMDLAPFMAQCQIF	120
Qy	180	MEKSTRDFSVDVRVRCVFTSMVTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFPDQR	239
Db	121	MEKSTRDFSVDVRVRCVFTSMVTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFPDQR	180
Qy	240	REVAKRIRRLROGMGSVIDYSNAFQIAQDLWNEPALIDQYHEGLSDHIOBELSHLEV	299
Db	181	REVAKRIRRLROGMGSVIDYSNAFQIAQDLWNEPALIDQYHEGLSDHIOBELSHLEV	240
Qy	300	AKSLSALIGOCIHIERRLARAAARPRSPRALVLPHTASHHQVDPTPEVGGARMRLTQ	359
Db	241	AKSLSALIGOCIHIERRLARAAARPRSPRALVLPHTASHHQVDPTPEVGGARMRLTQ	300
Qy	360	EKKERRKLNLCYCGTGGHYADNCAPAKASKSPAGNSAPL	401
Db	301	EKKERRKLNLCYCGTGGHYADNCAPAKASKSPAGNSAPL	342
RESULT	2		
ID	Q96A68	PRELIMINARY;	PRT; 325 AA.
AC	Q96A68;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	MEF3 LIKE 1 (PATERNALLY EXPRESSED GENE 10 ORF1).		
GN	MEF3L1 OR PEG10.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=TESTIS;		
RL	Nakamura Y., Furukawa Y.;		
RL	Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=TESTIS;		
RA	Sato S., Furukawa Y.;		
RT	"Isolation of MEF3 like gene 1.";		
RL	Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.		
RL	[3]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21218929; PubMed=11318613;		
RA	Ono R., Kobayashi S., Wagatsuma H., Aisaka K., Kohda T.,		
RA	Kaneko-Ishino T., Ishino F.;		
RT	"A Retrotransposon-Derived Gene, PEG10, Is a Novel Imprinted Gene		
RL	Located on Human Chromosome 7q21.";		
DR	Genomics 73:232-237(2001).		
DR	ENBL; AB049150; BAB68387.1; -		
DR	ENBL; AB049834; BAB43951.1; -		
SQ	SEQUENCE 325 AA; 36965 MW; 11864CFA97F2A76 CRC64;		
Query Match	79.6%;	Score 1703;	DB 4; Length 325;
Best Local Similarity	100.0%;	Pred. No. 3e-123;	
Matches	325; Conservative	0; Mismatches	0; Indels 0; Gaps
Qy	77	MTERRDELSEETNNLRKVMKQSENNLNQSQVKLTTEENTTLREQVEPTPEDEDDIE	136
Db	1	MTERRDELSEETNNLRKVMKQSENNLNQSQVKLTTEENTTLREQVEPTPEDEDDIE	60
Qy	137	LRGAAAAAAPPPIIEECPEDLPEKFDGNPMDLAPFMAQCQIFMEKSTRDFSVDVRVRCF	196
Db	61	LRGAAAAAAPPPIIEECPEDLPEKFDGNPMDLAPFMAQCQIFMEKSTRDFSVDVRVRCF	120
Qy	197	VTSMVTGRAARWASAKLERSHYLMHNYPAFMMEMKHVFPDQRREYAKRKIRLRQGMGS	256







Qy	120	LREQVEPTDEDDDIELRGAAAAAAPPPIIEECPEDLPKFDGPNPDLAPFMAQCQIF	179
Dd	73	LOGRAREAQELKEVATLRAAAANVISAPVBGRERKLINTPATDGTGGOLKGHLVQIRY	132
Qy	180	MEKSTRDFSVDVRVCVTSMMTGRAARW-----ASAKLESHYLHNVPAPM	227
Dd	133	QA FHMGTFQNDTERVVAHAFTLRGRALAWPEPQQEWLNPNVKEYSQEVNRNFFSFDGYV	192
Qy	228	MEMKHVPEDPORREAVAKKRIRRLROGMGSVIDYSNAFQMIAOQLDNNEPALIDQYHEGLS	287
Dd	193	KALQSLELDPEKQRQAEERDLSNLRNKSATL-YAAEFRRLLAAOLDMTDESKFAFYQGUK	251
Qy	288	DHIQEELSHELVAKSLSALIGQCIIH-----ERRLARAAAARPSPRALVLPHIASHH	342
Dd	252	DDVKDEMAKVDTTPSFLDYVEYAIKTDNRLFERRKEKEKRPUNSGRYQWQPQORFPNN	311
Qy	343	QVDPTEP-----VGGARMRLTQ-----BEKERKKLNILCYCGTG	377
Dd	312	QRNDNRPRNNWNSTQRSTSYGHSGPMPLDSATOKTSWRNNNQKNDNWKKDKOCFCNCOK	371
Qy	378	GHYADNCPA	386
Dd	372	GHLAYECPPA	380

```

RESULT      9
Q95KK1
ID          Q95KK1      PRELIMINARY;      PRT;      400 AA.
AC          Q95KK1;
DT          01-DEC-2001 (TrEMBLrel. 19, Created)
DT          01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT          01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE          HYPOTHETICAL 46.6 KDA PROTEIN.
OS          Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC          Cercopitheciinae; Macaca.
OX          NCBI_TaxID=9541;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          TISSUE=CEREBELLUM CORTEX;
RA          Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA          Suzuki Y., Sugano S., Hashimoto K.;
RT          "Isolation of full-length cDNA clones from macaque brain cDNA
RT          libraries.";
RL          Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR          EMBL; AB060816; BAB46855.1; -
KW          Hypothetical protein.
SQ          SEQUENCE      400 AA; 46612 MW; 911D7B5ADD0EB56 CRC64:

```

Query Match	10.4%	Score	223.5	DB	6	Length	400
Best Local Similarity	28.7%	Pred. No.	2.6e-09				
Matches	62	Conservative	39	Mismatches	84	Indels	31
Gaps							
QY	175	QCQIFMEKSTRDFSVDVRVCFVTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKHVF	234				
Db	2	QLETFIADHEDHFGGAEERVAFLISFTTGEAKDNAISVTOEGSPHANFPRLDEIRKEF	61				
QY	235	EDPORREVAKKKIRRLRQGMGSVIDYSNAFOMIAODLPWNEPALIDQYHEGLSDHQBEL	294				
Db	62	CGPIPPRVAKAIKRLKOGHTCLGSYADAFOFQAQFSWDCHLQNQLKLGLSEFFRKEL	121				
QY	295	SHLEVAKSLSAGCOTHIERRLARAAAARKPRPPP-----RALVLPHI-----	338				
Db	122	LWSTEMADLDLIEECVEIERKV-----RVPKPIPLPGCVNIIPLPFPAPNEESEEDEYY	175				
QY	339	-----ASHQVDPTEPVGGARMRLTQEEKERRK	367				
Db	176	SEDEDQACRHRLHPKDQRN--RWRAFQENKEKEE	209				

```

RESULT 10
Q9P2M8 PRELIMINARY; PRT; 1418 AA.
AC Q9P2M8:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1318 PROTEIN (FRAGMENT).
GN KIAA1318.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE=20161126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
DR EMBL; AB037739; BAA92556.1; -.
FT NON_TER 1
SQ SEQUENCE 1418 AA; 147474 MW; C3F6005F4F4374CD CRC64;
```

```

Query Match          9.0%;   Pred. 192.5;   DB 4;   Length 1418;
Best Local Similarity 20.4%;   Pred. NO. 3e-06;
Matches      78; Conservative 57; Mismatches 174; Indels 73; Gaps 10;

QY    25 PHRSEATAGRSPTTIVTLGDCPPPPPPPPNNNNNSKHTGHKSAC-----VPMTE 79
       |||::|||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    1048 PIRASAGARSTSFWRASVSSMPPLP-----RATSGCGMGMSMQMT- 1092
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY    80 RRRDELSEIINLRKVMMKQBENNLOSQVO-----KLTEENTTLRQVEPTPD 130
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    1093 -----ATDSRCGSTPLMRASPGGTMTSPQTAFGVMTSPETKATDGEASTSHINITASG 1146
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY    131 EDDDIELRGAAAAAAPPPIEECECPDJPEXFDGNPDLAPEMA----- 174
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    1147 SKPTSHMTATTETAKPP--KEVP-----SFGMLTALCYLLBEEQAARGSCSVE 1195
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY    175 -QCQIFMEKSTRDFSVDVRVCFTVSMTTGRAARWASAKLERSHYLMHNYPAFNMEMKHV 233
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    1196 EWEIDEEQMKGFLLDDSERMAFLVSLHLGAARWFILQMEVGEPLSHENKSFLLRSSQGI 1255
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY    234 FEDPORREVAKKIRRLROGGSGVDVYSNAQMIAODLDWNEPALIDQYHEGLSDHIQEE 293
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    1256 YDSLSEIDILSAVLCHPKOGOKSVROYATDFLLLARHLWSDAIIURFLGLESLSEAVTTK 1315
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY    294 LSH---LEVAKLSALIGOCIHERLARAANAARKPRSPRALYLPHASHHOVDTPVG 351
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    1316 MGRIFLKVAGSLKELIDRSLYTECOLAE-----EKDSPNGSSOQLPTACKRN----NEEAM 1367
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY    352 GARMLTOQEERRKKLNLCLY 373
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    1368 GNELSQQQOTEHQHVSKRCY 1389
       |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

RESULT	11	
ID	Q94115	PRELIMINARY; PRT; 1810 AA.
AC	Q94115;	
DT	01-DEC-2001	(TRENBLrel. 19, Created)
DT	01-DEC-2001	(TRENBLrel. 19, Last sequence update)
DT	01-DEC-2001	(TRENBLrel. 19, Last annotation update)
DE	PUTATIVE RETROELEMENT.	
OS	Oryza sativa (Rice).	
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta	
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
OC	Ehrhartoideae; Oryzeae; Oryza.	
OX	NCBI_TaxID=4530;	

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RT McCombie W.R.;
RL "Rice genomic sequence.";
RM Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Cordum H., Johnson D., Minx P., de la Bastide M., Nascimento L.U.,
RA Spiegel L.A., Kirchoff K.A., King L., Preston R.R., Vil M.D.,
RA Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA Rodriguez S., Cunniss D.M., Balija V.S., Shah R.S., Bahret A.,
RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA Rodriguez S., Cunniss D.M., Balija V.S., Shah R.S., Bahret A.,
RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
[4]
RN RP SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA Rodriguez S., Cunniss D.M., Balija V.S., Shah R.S., Bahret A.,
RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC022352; AAK51574.1; -.
SQ SEQUENCE 1810 AA; 205746 MW; 5AB92CD640BAB1F2 CRC64;

Query Match 8.8%; Score 189; DB 10; Length 1810;
Best Local Similarity 20.7%; Pred. No. 7.6e-06;
Matches 100; Conservative 57; Mismatches 125; Indels 202; Gaps 24;

QY 5 RVLTKKR-RSGRGQDPLHPRSEATAGRSPTPTVTGLGDCPPPPPPPPPPNNNNN 63
DB 304 QMVTTRNVNTGEGNPEGSNPHQ-----GNPPPP-----PPPPPPPPDTN---- 346
QY 64 SKHTGHSACVPMNTERRRDELSEENLNREKVMKQSENNLNQSQVOKLTEENTTLREQ 123
DB 347 -----ALTOI-----LAQANM-----TAFLLH 366
QY 124 VEPTPEDDDDIELRGAAAAAAPPPIEECPEDL---PEKF--DGNP-DMLAPPMA--- 174
DB 367 LONPPON-----APPPPHSKLAELFLRIRPPTFSSNNPVDALDHLHVGK 414
QY 175 -----QCQIFMEKSTRDFSVDVRVCFVTSMTGRAARW-----ASA 211
DB 415 KLDTVQCS-----DEEKVFAAHQLOGPASPALMWDHFOATQPEGQPTWTRFTA 462
QY 212 KLSRHYLMHNYPAFMEMKHFVEDPQRRVAKRKIRRLROGMSGVIDYSNAFOMIA--- 268
DB 463 AFRTH-----VPAGVVALK-----KREFRELKQGNRSWMEYLHEFNNLARYA 505
QY 269 -QLDWNNEPALIDQYHEGLSDHIOEELSHLEVA---KSLSALIGQCIIHIE----- 314
DB 506 PEDVREDE-----EKQEKFLAGMDPELSVRLVSGDYPDFQRLVDKSIIRLEAKHKELES 560
QY 315 RLIA-----RAAARPRSPRALVPLPHIASHHOVD----- 345
DB 561 RLANFRNQOGANQVRVYTNYPYGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 620
QY 346 -----PT-EPVGGARMRLTOEKEKRRRKLNLCLYCGTGGHYADNCP---AKAKSSP 393
DB 621 TRAPRPTPTVQVQGR-----DAQOQRLCFCNCFEPGHFADKCPKPRQOQAPP 672
```

```

QY 394 AGNS 397
DB 673 RSNN 676

RESULT 12
Q94DA0 PRELIMINARY; PRT; 1500 AA.
AC Q94DA0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE POLYPROTEIN.
GN P0439E11.5.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0439E11.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP003315; BAB61255.1; -.
KW Polyprotein.
SQ SEQUENCE 1500 AA; 171517 MW; 5DE27E21F9ABE861 CRC64;

Query Match 8.6%; Score 185; DB 10; Length 1500;
Best Local Similarity 20.3%; Pred. No. 1.2e-05;
Matches 98; Conservative 53; Mismatches 125; Indels 206; Gaps 21;

QY 6 VLTKKR-RSGRGQDPLHPRSEATAGRSPTPTVTGLGDCPPPPPPPPPPNNNNNS 64
DB 1 MVSTRNVNTGEGNPEGSNPN-----POGNPPPPPPPPDTN----- 37
QY 65 KHTGHSACVPMNTERRRDELSEENLNREKVMKQSENNLNQSQVOKLTEENTTLREQ 124
DB 38 -----ALTOIQAQANMNAFLHHLQNLPOQ----- 65
QY 125 EPTPEDDDDIELRGAAAAAAPPPIEECPEDL---PEKF--DGNP-DMLAPPMA--- 174
DB 66 -----APPPPHSKLAELFLRIRPPTFSSNNPVDALDHLHVGK 106
QY 175 -----QCQIFMEKSTRDFSVDVRVCFVTSMTGRAARW-----ASAK 212
DB 107 LDTVQCS-----DEEKVFAAHQLOGPASPALMWDHFOATQPEGQPTWTRFTA 154
QY 213 LERSHYLMHNYPAFMEMKHFVEDPQRRVAKRKIRRLROGMSGVIDYSNAFOMIA--- 268
DB 155 FRTH-----VPAGVVALK-----KREFRELKQGNRSWMEYLHEFNNLARYA 197
QY 269 -QLDWNNEPALIDQYHEGLSDHIOEELSHLEVA---KSLSALIGQCIIHIE-----R 315
DB 198 EDVREDE-----EKQEKFLAGMDPELSVRLVSGDYPDFQRLVDKSIIRLEAKHKELES 252
QY 316 RLIA-----RAAARPRSPRALVPLPHIASHHOVD----- 345
DB 253 RLANFRNQOGANQVRVYTNYPYGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 312
QY 346 -----PT-EPVGGARMRLTOEKEKRRRKLNLCLYCGTGGHYADNCP---AKAKSSP 395
DB 313 APRPPTPTVQVQGR-----DAQOQRLCFCNCFEPGHFADKCPKPRQOQAPP 364
QY 396 NS 397
DB 365 NN 366

RESULT 13
```

```
Q94I04
ID Q94I04 PRELIMINARY; PRT; 1862 AA.
AC Q94I04;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE RETROBLENT.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RC STRAIN=NIPPONBARE;
RC STRAIN=NIPPONBARE;
RC Cordon H., Johnson D., Minx P., de la Bastide M., Nascimento L.U.,
RA Spiegel L.A., Kirchoff K.A., King L., Preston R.R., Vil M.D.,
RA Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA Rodriguez S., Cunnius D.M., Balija V.S., Shah R.S., Bahret A., Bal H.,
RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RC Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA Rodriguez S., Cunnius D.M., Balija V.S., Shah R.S., Bahret A.,
RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RA Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RC Palmer L.E., Bal H., Cordum H., Johnson D., Minx P., de la Bastide M.,
RA Nascimento L.U., Spiegel L.A., Kirchoff K.A., King L., Preston R.R.,
RA Vil M.D., Baker J.P., Miller B., Santos L., Zutavern T., Kuit K.H.,
RA Rodriguez S., Cunnius D.M., Balija V.S., Shah R.S., Bahret A.,
RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;
RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AC022352; AAK51585.1; -
SQ SEQUENCE 1862 AA; 211396 MW; 291B106D83539AB4 CRC64;

Query Match 8.6%; Score 185; DB 10; Length 1862;
Best Local Similarity 20.2%; Pred. No. 1.6e-05;
Matches 97; Conservative 56; Mismatches 133; Indels 194; Gaps 22;

QY 5 RVLUKTKRRSGRGQDPLGPHRSEATAGRSPTPTVTGLGDCPPPPPPPPNNNNNNNS 64
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
356 RVMVTRRNTAGDGNQPECSNNHQG-----NPPPP-----PPPPPPPPPDNTN----- 398
QY 65 KHTGKHSACVPNMTERRRDELSEINNLREKVMKQSENNNLSQVOKLTTEENTTLREQV 124
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
399 -----AILTQI-----LAQOANMM-----TAFUHLH 419
QY 125 EPTPEDEDDIELRGAAAAAPPPPIEECEPDL---PEKF--DGNP-DMLAPFMA--- 174
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
420 QNPPQON-----APPPPPQHSKLAELRIRPTFFSSNNPNVDALDHLHVGKK 467
QY 175 -----QCOIFMEKSTRDFSVDRVRCVFTSMMTGRAARW-----ASAK 212
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
468 LGTVQCS-----DEEKVIFAHHQLOGPASLMDHDFQATQPEGQITWARFTAA 515
QY 213 LERSHYLMHNPAPFMMEMKHVFEDPQREVAKRKIRLRQGMGSMVIDYSNAPOMIA---- 268
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
516 FRTH-----VPAGVALK-----KREFRELKQGNRSVMYELHFEFNRLARYAP 558
```

```
QY 269 QDLWNEPALIDQYHEGLSDHIOEELSHLEVA---KSLSALIGQCIHIE-----R 315
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
559 EDVREDE-----EKQKFLAGMDPELSVRLISGDYPDFQRLVDSIRLEAKHKELESKR 613
QY 316 RLA-----RAAAARKSRPPRALVLPHTASHHQVD----- 345
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
614 RLANFRNQGANQVRVRYTHPYPGSSSQOQQOQPSAPRPQFVVRVPQPOQOQOQGT 673
QY 346 -----PTEPVGGARWLTQEKERRRKLNLCLYCGTGCHYADNCP---AKAKSSPAGNS 397
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
674 RAPRPPTPTVOPGGRRDAQGQR-----LCFNCFEFGHFADKCPKRRQOQOQAPPRSN 728

RESULT 14
Q9AYCO PRELIMINARY; PRT; 1781 AA.
ID Q9AYCO;
AC Q9AYCO;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POLYPROTEIN.
GN OSJNEA0094H10.15.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoidae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=NIPPONBARE;
RA Spiegel L.A., Nascimento L.U., de la Bastide M., Kirchoff K.A.,
RA King L., Preston R.R., Vil M.D., Baker J.P., Miller B., Zutavern T.,
RA Rodriguez S., Santos L., Kuit K.H., Cunnius D.M., Balija V.S.,
RA Shah R.S., Bahret A., Bal H.P., O'Shaughnessy A., Dedhia N.N.,
RA McCombie W.R.;
RA "Genomic Sequence For Oryza sativa, Nipponbare Strain, Chromosome X,
RT Clone OSJNBa0058E19, complete sequence.";
RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AC080019; AAK13118.1; -
DR InterPro: IPR001969; Asp_protease.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00078; rvt; 1.
DR Pfam: PF00098; zf-CCHC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
KW RNA-directed DNA polymerase; Zinc-finger.
SQ SEQUENCE 1781 AA; 202845 MW; 9984A707A125FE8A CRC64;

Query Match 8.6%; Score 184.5; DB 10; Length 1781;
Best Local Similarity 20.1%; Pred. No. 1.6e-05;
Matches 95; Conservative 56; Mismatches 141; Indels 181; Gaps 19;

QY 5 RVLUKTKRRSGRGQDPLGPHRSEATAGRSPTPTVTGLGDCPPPPPPPPNNNNNNNS 64
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
285 RVMVTRRNTAGDGNQPECSNNHHG-----SPPPP-----PPPPPPPPPDNTN----- 327
QY 65 KHTGKHSACVPNMTERRRDELSEINNLREKVMKQSENNNLSQVOKLTTEENTTLREQV 124
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
328 -----AILTQIILAQOANMMTAFUHLH-----QNPPQHN----- 355
QY 125 EPTPEDEDDIELRGAAAAAPPPPIEECEPDL---PEKF--DGNP-DMLAPFMA--- 174
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
356 -----APPPPPQHSKLAELRIRPTFFSSNNPNVDALDHLHVGKK 396
QY 175 -----QCOIFMEKSTRDFSVDRVRCVFTSMMTGRAARW-----ASAK 212
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
397 LDTVQCS-----DEEKVIFAHHQLOGPASLMDHDFQATQPEGQITWARFTAA 444
QY 213 LERSHYLMHNPAPFMMEMKHVFEDPQREVAKRKIRLRQGMGSMVIDYSNAPOMIA---- 268
```



---

**THIS PAGE BLANK (USPTO)**







PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.

PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.5%; Score 10; DB 21; Length 174;  
Best Local Similarity 100.0%; Pred. No. 0.4;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 pppppppppn 57  
          |          |  
Db 18 pppppppppn 27

RESULT 3  
AAG18589  
ID AAG18589 standard; Protein; 209 AA.  
XX  
AC AAG18589;  
XX  
DT 17-Oct-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 20060.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137724.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.5%; Score 10; DB 21; Length 209;

Best Local Similarity 100.0%; Pred. No. 0.47;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 pppppppppn 57

Db 53 pppppppppn 62

RESULT 4

AAG11272  
ID AAG11272 standard; Protein; 235 AA.

XX

AC AAG11272;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 9926.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

OS

XX

PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.5%; Score 10; DB 21; Length 235;

Best Local Similarity 100.0%; Pred. No. 0.52;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 pppppppppn 57

Db 79 pppppppppn 88

RESULT 5

AAG37672

ID AAG37672 standard; Protein; 235 AA.

XX AAG37672;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 46360.

DE Arabidopsis thaliana.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.

[illegible]

PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 16-AUG-1999;	99US-0148684.
PR 17-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 21-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.
PR 22-OCT-1999;	99US-0160981.
PR 22-OCT-1999;	99US-0160989.
PR 25-OCT-1999;	99US-0161404.
PR 25-OCT-1999;	99US-0161405.
PR 25-OCT-1999;	99US-0161406.
PR 26-OCT-1999;	99US-0161359.
PR 26-OCT-1999;	99US-0161360.
PR 26-OCT-1999;	99US-0161361.
PR 28-OCT-1999;	99US-0161920.
PR 28-OCT-1999;	99US-0161992.
PR 28-OCT-1999;	99US-0161993.
PR 29-OCT-1999;	99US-0162142.
Query Match 2.5%; Score 10; DB 21; Length 237;	
Best Local Similarity 100.0%; Pred. No. 0.52;	
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 48 pppppppppn 57	
Db 81 pppppppppn 90	
RESULT 7	
AAG37671	
ID AAG37671 standard; Protein; 237 AA.	
XX	
AC AAG37671;	
XX	
DT 18-OCT-2000 (first entry)	
XX	
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46359.	
XX	
KW Protein identification; signal transduction pathway; metabolic pathway;	
KW hybridisation assay; genetic mapping; gene expression control; promoter;	
KW termination sequence.	
XX	
OS Arabidopsis thaliana.	
XX	
PN EP1033405-A2.	
XX	
PD 06-SEP-2000.	
XX	
PF 25-FEB-2000; 2000EP-0301439.	
XX	
PR 25-FEB-1999;	99US-0121825.
PR 05-MAR-1999;	99US-0123180.
PR 09-MAR-1999;	99US-0123548.
PR 23-MAR-1999;	99US-0125788.
PR 25-MAR-1999;	99US-0126264.
PR 29-MAR-1999;	99US-0126785.
PR 01-APR-1999;	99US-0127462.
PR 06-APR-1999;	99US-0128234.
PR 08-APR-1999;	99US-0128714.
PR 16-APR-1999;	99US-0129845.
PR 19-APR-1999;	99US-0130077.
PR 21-APR-1999;	99US-0130449.
PR 23-APR-1999;	99US-0130510.
PR 28-APR-1999;	99US-0130891.
PR 28-APR-1999;	99US-0131449.

PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.



```
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      2.5%; Score 10; DB 21; Length 237;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPN 57
Db 81 PPPPPPPPN 90

RESULT 8
AAG39269
ID AAG39269 standard; Protein; 237 AA.
XX
AC AAG39269;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48563.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
```

PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 2.5%; Score 10; DB 21; Length 237;  
Best Local Similarity 100.0%; Pred. No. 0.52; Mismatches 0; Indels 0; Gaps 0;  
Matches 10; Conservative 0;  
  
QY 48 PPPPPPPPPN 57  
Db 81 PPPPPPPPPN 90  
  
RESULT 9  
AAG11270  
ID AAG11270 standard; Protein; 240 AA.  
XX  
AC AAG11270;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9924.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0123786.  
PR 29-MAR-1999; 99US-0126264.  
PR 01-APR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 08-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 05-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

```
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      2.5%; Score 10; DB 21; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 pppppppppn 57
Db 84 pppppppppn 93

RESULT 10
RAG37670
ID AAG37670 standard; Protein; 240 AA.
XX
AC AAG37670;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46358.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 200EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
```

Best Local Similarity 100.0%; Pred. No. 0.53;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPPN 57  
|||||  
Db 84 PPPPPPPPPN 93

RESULT 11  
AAG39268  
ID AAG39268 standard; Protein; 240 AA.  
AC AAG39268;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48562.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.5%; Score 10; DB 21; Length 240;  
Best Local Similarity 100.0%; Pred. No. 0.53;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 pppppppppn 57

Db 84 pppppppppn 93

RESULT 12  
AAG39267  
ID AAG39267 standard; Protein; 263 AA.  
XX  
AC AAG39267;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48561.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0148171.  
PR 10-AUG-1999; 99US-0148319.  
PR 11-AUG-1999; 99US-0148341.  
PR 12-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 03-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 2.5%; Score 10; DB 21; Length 263;  
Best Local Similarity 100.0%; Pred. No. 0.57; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 pppppppppn 57  
|||||||  
Db 107 pppppppppn 116

RESULT 13  
ABG25300  
ID ABG25300 standard; Protein; 697 AA.  
XX  
AC ABG25300;  
XX

DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #25291.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
PF 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR N-PSDB; AAS89487.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 20; SEQ ID No 55659; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 697 AA;

Query Match 2.5%; Score 10; DB 22; Length 697;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 AAAAAAPPPP 149  
|||||||  
Db 520 aaaaaapppp 529

RESULT 14  
AAP20304  
ID AAP20304 standard; peptide; 9 AA.  
XX  
AC AAP20304;  
XX  
DT 30-NOV-1992 (first entry)



XX DE Peptide for use in an anti-wrinkle cosmetic.  
XX KW Cream; non-irritating.  
XX XX Synthetic.  
XX PN JP57002214-A.  
XX PD 07-JAN-1982.  
XX PF 06-JUN-1980; 80JP-0076267.  
XX PR 05-JUN-1980; 80JP-0076268.  
XX PA (MTU ) MITSUBISHI CHEM IND KK.  
XX PA (KANE-) KANEBO KESHOIN KK.  
XX DR WPI; 1982-12597E/07 (12597E).  
XX PT Peptide-contg. cosmetic - preventing formation of fine wrinkles  
XX PS Claim 1; Page 4; 6pp; Japanese.  
XX CC The peptide is used in a cosmetic to prevent the formation of fine  
CC wrinkles without irritating the skin. The peptide has at its N-  
CC terminus an organic gp. which is capable of reacting with hydrogen  
CC or amino gps. e.g. a benzoyl, acetyl, formyl, tosyl gp. etc. The  
CC peptide may be prepd. by standard peptide synthesis techniques.  
XX SQ Sequence 9 AA;  
  
Query Match 2.2%; Score 9; DB 3; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.4e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 48 pppppppppp 56  
Db 1 pppppppppp 9  
  
RESULT 15  
AAU09969  
ID AAU09969 standard; peptide; 10 AA.  
XX AC AAU09969;  
XX DT 16-JAN-2002 (first entry)  
XX DE Proline linker sequence.  
XX KW Cancer; caspase; apoptosis; Alzheimer's disease;  
KW Down's syndrome; prion protein; Bovine Spongiform Encephalopathy;  
KW BSE; acquired immune deficiency syndrome; AIDS; proline linker;  
KW Creutzfeld-Jacob disease; CJD.  
XX OS Synthetic.  
XX PN WO200175453-A2.  
XX PD 11-OCT-2001.  
XX PF 04-APR-2001; 2001WO-GB01540.  
XX PR 04-APR-2000; 2000GB-0008254.  
XX PR 04-APR-2000; 2000GB-0008256.  
XX PA (MEDI-) MEDICAL RES COUNCIL.  
XX PI Tse E, Rabbitts T;  
XX DR WPI; 2001-626485/72.

XX PT Inducing cell to generate detectable signal and detecting entity within  
PT cell, comprises allowing stable interaction of first reporter and  
PT second reporter by binding with the entity and generating signal -  
XX PS Disclosure; Page 45; 76pp; English.  
XX CC This sequence represents a proline linker sequence used for cloning  
CC F-box proteins from mammalian cells or tissues in the method of  
CC the invention. The invention comprises a method of inducing a cell to  
CC generate a detectable signal and detecting an entity within a cell. The  
CC method involves providing a cell containing an entity with a first and  
CC second reporter where a stable interaction between the reporters  
CC leads to generation of a detectable signal which can be monitored. Also  
CC included is a method for killing a cell in which a cell containing a  
CC given entity is given a first reporter containing a first immunoglobulin  
CC (Igi) and a second reporter containing a second immunoglobulin (Ig2).  
CC The Igi molecule is linked to a first caspase molecule and a second  
CC caspase molecule is linked to Ig2, where the two caspases are capable of  
CC stably interacting to generate caspase activity and induce apoptosis.  
CC This method can also be used for destruction of a polypeptide in a cell,  
CC by providing a cell containing a polypeptide with two reporters where a  
CC stable interaction between these reporters leads to generation of  
CC protease activity, and proteolysis of the polypeptide. The function of  
CC unknown genes can be determined with this method by using this technique  
CC to destroy the polypeptide encoded by the gene and observing a  
CC phenotype. These methods are useful for detecting an entity within a cell  
CC (e.g. senile plaques in Alzheimer's disease, neurofibrillary tangles in  
CC Down's syndrome or infectious prion proteins associated with Bovine  
CC Spongiform Encephalopathy (BSE), Creutzfeld-Jacob disease (CJD) or  
CC proteins associated with acquired immune deficiency syndrome (AIDS)),  
CC killing a cell, destruction of a polypeptide in a cell, functional  
CC genomic studies and in the treatment or diagnosis of a cancer.  
XX SQ Sequence 10 AA;  
  
Query Match 2.2%; Score 9; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 48 pppppppppp 56  
Db 1 pppppppppp 9  
  
Search completed: August 13, 2002, 15:13:36  
Job time: 142 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:12:04 ; Search time 20.85 Seconds  
(without alignments)  
1848.049 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 401  
Sequence: 1 MNKRVLTKKRRSGGQD.....DNCPRKASKSPAGNPAPL 401

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	15	3-7	232	2 JE0163	myelin expression
2	10	2-5	250	2 T16342	hypothetical prote
3	10	2-5	494	2 B96534	hypothetical prote
4	10	2-5	1018	2 T43168	hypothetical prote
5	10	2-5	1259	2 T16038	hypothetical prote
6	9	2-2	90	2 T18068	hypothetical prote
7	9	2-2	135	2 E96750	hypothetical prote
8	9	2-2	141	2 A34043	hypothetical proli
9	9	2-2	142	2 B41132	collagen-related p
10	9	2-2	148	2 S39206	proline-rich prote
11	9	2-2	149	2 A41132	collagen-related p
12	9	2-2	161	2 H85120	probable proline-r
13	9	2-2	165	2 C71717	hypothetical prote
14	9	2-2	172	2 D41132	collagen-related p
15	9	2-2	172	2 T27505	hypothetical prote
16	9	2-2	182	2 T30760	hypothetical prote
17	9	2-2	186	2 C41132	collagen-related p
18	9	2-2	195	2 T07735	nodulin-20a - soyb
19	9	2-2	219	2 T10563	proline-rich prote
20	9	2-2	220	2 S52005	nodulin-30 (Npv30)
21	9	2-2	221	2 H82857	tonB protein XF000
22	9	2-2	229	1 W4WLB	E4 protein - human
23	9	2-2	236	2 T02577	probable AP2 domai
24	9	2-2	246	2 T46446	hypothetical prote
25	9	2-2	250	1 A31757	homeotic protein H
26	9	2-2	250	2 T08908	hypothetical prote
27	9	2-2	251	1 B60492	homeotic protein H
28	9	2-2	255	2 C36222	phosphoprotein pho
29	9	2-2	257	2 T00406	hypothetical prote

ALIGNMENTS

RESULT 1

JE0163  
myelin expression factor-3 - mouse  
N:Alternate names: MYEF-3  
C:Species: Mus musculus (house mouse)  
C>Date: 03-Jul-1998 #sequence\_revision 10-Jul-1998 #text\_change 07-May-1999  
C:Accession: JE0163  
R:Stepiowski, A.; Krynska, B.; Tretiakova, A.; Haas, S.; Khailli, K.; Amini, S.  
Biochem. Biophys. Res. Commun. 243, 295-301, 1998  
A:Title: MYEF-3, a developmentally controlled brain-derived nuclear protein which spe  
A:Reference number: JE0163; MUID:98139908  
A:Accession: JE0163  
A:Molecule type: mRNA  
A:Residues: 1-232 <STE>  
A:Experimental source: brain  
C:Keywords: phosphoprotein  
F:40-60/Domain: transmembrane #status predicted <TM>  
F:31,175/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predi  
F:31,142,185,201/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #stat

Query Match 3.7%; Score 15; DB 2; Length 232;  
Best Local Similarity 100.0%; Pred. No. 7.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 259 DYSNAFQMIQDLQDW 273  
|||||  
Db 106 DYSNAFQMIQDLQDW 120

RESULT 2

T16342  
hypothetical protein F42C5.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T16342  
R:Du, Z.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid F42C5.  
A:Reference number: Z18497  
A:Accession: T16342  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-250 <DUZ>  
A:Cross-references: EMBL:U040799; NID:g1065935; PID:g1065941; PIDN:AAA81484.1; CESP:F4  
C:Genetics:  
A:Gene: CESP:F42C5.7  
A:Introns: 15/2; 40/3; 167/3; 191/3

Query Match 2.5%; Score 10; DB 2; Length 250;  
Best Local Similarity 100.0%; Pred. No. 0.29;

```

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 C P P P P P P P P P 56
      |||||
Db 102 C P P P P P P P P P 111

RESULT 3
B96534
hypothetical protein F14J22.4 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B96534
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, O.; Alonso,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B96534
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <STO>
A:Cross-references: GB:AE005173; NID:g10120434; PIDN:AGL13059.1; GSPDB:GN00141
A:Gene: F14J22.4
A:Map position: 1
C:Genetics:

Query Match 2.5%; Score 10; DB 2; Length 494;
Best Local Similarity 100.0%; Pred. No. 0.52;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 D C P P P P P P P P 55
      |||||
Db 60 D C P P P P P P P P 69

RESULT 4
T43168
hypothetical protein cyk-1 - Caenorhabditis elegans (fragment)
C:Species: Caenorhabditis elegans
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T43168
R:Swan, K.A.; Severson, A.F.; Carter, J.C.; Martin, P.R.; Schnabel, H.; Schnabel, R.; Bc
J. Cell Sci. 111, 2017-2027, 1998
A:Title: cyk-1: a C. elegans FH gene required for a late step in embryonic cytokinesis.
A:Reference number: 222311; MUID:98311662
A:Accession: T43168
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1018 <SWA>
A:Cross-references: EMBL:AF062008; PIDN:AAC17501.1
C:Genetics:
A:Gene: cyk-1
A:Map position: 3

Query Match 2.5%; Score 10; DB 2; Length 1018;
Best Local Similarity 100.0%; Pred. No. 0.97;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 C P P P P P P P P P 56
      |||||
Db 351 C P P P P P P P P P 360

RESULT 5

```

```

T16038
hypothetical protein F11H8.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16038
R:Menezes, S.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F11H8.
A:Reference number: Z18450
A:Accession: T16038
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1259 <MEN>
A:Cross-references: EMBL:U40187; NID:g1055194; PID:g1055198; PIDN:AAA81161.1; CESP:F1
C:Genetics:
A:Gene: CESP:F11H8.4
A:Introns: 160/3; 192/1; 212/3; 291/3; 350/3; 417/2; 511/3; 632/1; 677/3; 895/1; 1073

Query Match 2.5%; Score 10; DB 2; Length 1259;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 C P P P P P P P P P 56
      |||||
Db 768 C P P P P P P P P P 777

RESULT 6
T18068
hypothetical protein a566L - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18068
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL data Library, May 1999
A:Reference number: Z18806
A:Accession: T18068
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-90 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC97003.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: a566L

Query Match 2.2%; Score 9; DB 2; Length 90;
Best Local Similarity 100.0%; Pred. No. 0.98;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 P P P P P P P P P P 56
      |||||
Db 39 P P P P P P P P P P 47

RESULT 7
E96750
hypothetical protein F28P22.21 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E96750
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719

```

A;Accession: E96750  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-135 <STO>  
A;Cross-references: GB:AE0051173; NID:96648168; PIDN:AAF21168.1; GSPDB:GN00141  
C;Genetics:  
A;Map position: 1

Query Match 2.2%; Score 9; DB 2; Length 135;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 38 PPPPPPPPP 46

RESULT 8  
A34043  
hypothetical proline-rich protein 1 - polychaete (Owenia fusiformis) (fragment)  
C;Species: Owenia fusiformis  
C;Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 01-Dec-2000  
C;Accession: A34043; B34043  
R;Bakalara, N.; Collet, J.; Planells, R.; Thouveny, Y.; Fontes, M.  
Biochem. Biophys. Res. Commun. 166, 66-73, 1990  
A;Title: Presence in invertebrate genomes of sequences characterized by the repetition  
A;Reference number: A90159; MUID:90147742  
A;Accession: A34043  
A;Molecule type: DNA  
A;Residues: 1-141 <BAK>  
A;Cross-references: GB:M32217  
A;Accession: B34043  
A;Molecule type: DNA  
A;Residues: 59-136 <BA2>  
A;Cross-references: GB:M32217

Query Match 2.2%; Score 9; DB 2; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 9 PPPPPPPPP 17

RESULT 9  
B41132  
collagen-related protein 2 - Hydra magnipapillata (fragment)  
C;Species: Hydra magnipapillata  
C;Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 01-Dec-2000  
C;Accession: B41132; S21930  
R;Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.  
J. Cell Biol. 115, 1159-1169, 1991  
A;Title: Mini-collagens in hydra nematocytes.  
A;Reference number: A41132; MUID:92064646  
A;Accession: B41132  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-142 <KUR>  
A;Cross-references: EMBL:X61046; NID:99448; PIDN:CAA43380.1; PID:99449  
A;Note: submitted to the EMBL Data Library, July 1991  
C;Superfamily: unassigned collagens

Query Match 2.2%; Score 9; DB 2; Length 142;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 46 PPPPPPPPP 54

RESULT 10

S39206  
proline-rich protein V-beta 1 precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 19-May-2000  
C;Accession: A53118; S39206  
R;Court, Y.; Rosinski-Chupin, I.; Rougeon, F.  
J. Biol. Chem. 269, 520-527, 1994  
A;Title: A new proline-rich protein precursor expressed in the salivary glands of the  
A;Reference number: A53118; MUID:94103265  
A;Accession: A53118  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-148 <COU>  
A;Cross-references: EMBL:X74229; NID:9433616; PIDN:CAA52300.1; PID:9433617  
A;Note: submitted to the EMBL Data Library, July 1993  
C;Genetics:  
A;Gene: VCS-beta1  
A;Introns: 18/3  
C;Superfamily: proline-rich peptide P-B  
C;Keywords: glycoprotein  
F;1-18/Domain: signal sequence #status predicted <SIG>  
F;19-148/Product: proline-rich protein V-beta 1 #status predicted <WAT>  
F;133,143/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 2.2%; Score 9; DB 2; Length 148;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 35 PPPPPPPPP 43

RESULT 11

A41132  
collagen-related protein 1 precursor - Hydra magnipapillata  
C;Species: Hydra magnipapillata  
C;Date: 05-Jun-1992 #sequence\_revision 05-Jun-1992 #text\_change 01-Dec-2000  
C;Accession: A41132; S21929  
R;Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.  
J. Cell Biol. 115, 1159-1169, 1991  
A;Title: Mini-collagens in hydra nematocytes.  
A;Reference number: A41132; MUID:92064646  
A;Accession: A41132  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-149 <KUR>  
A;Cross-references: EMBL:X61045; NID:99446; PIDN:CAA43379.1; PID:99447  
A;Note: submitted to the EMBL Data Library, July 1991  
C;Superfamily: unassigned collagens

Query Match 2.2%; Score 9; DB 2; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||

Db 52 PPPPPPPPP 60

RESULT 12

H85120  
probable proline-rich protein [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C;Accession: H85120  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: H85120  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-161 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267842; PIDN:CAB81244.1; GSPDB:GN00140  
C:Genetics:  
A:Map position: 4

Query Match 2.2%; Score 9; DB 2; Length 161;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||  
Db 120 PPPPPPPPP 128

RESULT 13  
C71717  
hypothetical protein RP084 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000  
C:Accession: C71717  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499  
A:Accession: C71717  
A>Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-165 <AND>  
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14554.1; PID:el34239  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: RP084

Query Match 2.2%; Score 9; DB 2; Length 165;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||  
Db 36 PPPPPPPPP 44

RESULT 14  
D41132  
collagen-related protein 4 - Hydra magnipapillata (fragment)  
C:Species: Hydra magnipapillata  
C:Date: 05-Jun-1992 #sequence\_revision 12-Jun-1992 #text\_change 01-Dec-2000  
C:Accession: D41132; S21932  
R:Kurz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.  
J. Cell Biol. 115, 1159-1169, 1991  
A:Title: Mini-collagens in hydra nematocytes.  
A:Reference number: A41132; MUID:92064646  
A:Accession: D41132  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-172 <KUR>  
A:Cross-references: EMBL:X61048; NID:g9452; PIDN:CAA43382.1; PID:g9453  
A:Note: the authors translated the codon GGT for residue 142 as Pro, and TTA for residue  
A:Note: submitted to the EMBL Data Library, July 1991  
C:Superfamily: unassigned collagens

Query Match 2.2%; Score 9; DB 2; Length 172;  
Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
|||||  
Db 51 PPPPPPPPP 59

RESULT 15  
T27505  
hypothetical protein ZC168.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27505  
R:Berks, M.  
submitted to the EMBL Data Library, March 1996  
A:Reference number: Z20378  
A:Accession: T27505  
A>Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-172 <WIL>  
A:Cross-references: EMBL:Z70312; PIDN:CAA94385.1; GSPDB:GN00022; CESP:ZC168.5  
A:Experimental source: clone ZC168  
C:Genetics:  
A:Gene: CESP:ZC168.5  
A:Map position: 4  
A:Introns: 112/3

Query Match 2.2%; Score 9; DB 2; Length 172;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 CPPPPPPPP 55  
|||||  
Db 38 CPPPPPPPP 46

Search completed: August 13, 2002, 15:14:26  
Job time: 142 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:13:04 ; Search time 13.48 Seconds  
(without alignments)  
1151.820 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 401  
Sequence: 1 MRNRVLTKRRSRGRGQD.....DNCAPAKASKSPAGNSAPL 401

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.5	387	1 UL33_RCMVM	O12000 rat cytochrome
2	9	2.2	141	1 YPRO_OWEFU	P21260 owenia fusi
3	9	2.2	165	1 Y084_RICPR	Q9ze63 rickettsia
4	9	2.2	223	1 TONB_XANCP	O34261 xanthomonas
5	9	2.2	229	1 VE4_HPV08	P06425 human papil
6	9	2.2	250	1 HX84_MOUSE	P10284 mus musculu
7	9	2.2	251	1 HX84_HUMAN	P17483 homo sapien
8	9	2.2	272	1 TRPC_MYCTU	O06129 mycobacteri
9	9	2.2	281	1 FASL_HUMAN	P48023 homo sapien
10	9	2.2	287	1 SMN_BOVIN	O18870 bos taurus
11	9	2.2	289	1 SMN_RAT	O35876 rattus norv
12	9	2.2	290	1 HX88_HUMAN	P13378 homo sapien
13	9	2.2	294	1 SMN_HUMAN	O16637 homo sapien
14	9	2.2	297	1 HX1M_CHICK	O93367 gallus gall
15	9	2.2	309	1 HXA4_CHICK	P17277 gallus gall
16	9	2.2	345	1 CEBB_HUMAN	P17676 homo sapien
17	9	2.2	348	1 CEBB_BOVIN	O02755 bos taurus
18	9	2.2	353	1 CEBB_BOVIN	O02754 bos taurus
19	9	2.2	367	1 P53_CHICK	P10360 gallus gall
20	9	2.2	377	1 PFTA_MOUSE	O61239 mus musculu
21	9	2.2	377	1 PFTA_RAT	O04631 rattus norv
22	9	2.2	389	1 FL_ORYSA	O24175 oryza sativ
23	9	2.2	389	1 NDPP_MOUSE	O03173 mus musculu
24	9	2.2	401	1 HME1_MOUSE	P09065 mus musculu
25	9	2.2	410	1 HXA3_HETFR	O91a21 heterodontu
26	9	2.2	415	1 ACRO_PIG	P08001 sus scrofa
27	9	2.2	421	1 ACRO_HUMAN	P10323 homo sapien
28	9	2.2	431	1 ACRO_RABIT	P48038 oryctolagus
29	9	2.2	431	1 SMA6_CHICK	O9w734 gallus gall
30	9	2.2	436	1 GDF6_BOVIN	P55106 bos taurus
31	9	2.2	440	1 FXGA_CHICK	O98937 gallus gall
32	9	2.2	440	1 G3PT_MOUSE	O64467 mus musculu
33	9	2.2	449	1 WT1_HUMAN	P19544 homo sapien

34	9	2.2	449	1 WTL_PIG	O62651 sus scrofa
35	9	2.2	450	1 CYSP_TRYBB	P14658 trypanosoma
36	9	2.2	456	1 FXD1_MOUSE	Q01345 mus musculu
37	9	2.2	474	1 VP61_NPVOP	O10270 orgyia pseu
38	9	2.2	477	1 CAP2_RAT	P52481 rattus norv
39	9	2.2	485	1 SSGP_VOLCA	P21977 volvox cart
40	9	2.2	487	1 EBN2_EBV	P12978 epstein-bar
41	9	2.2	488	1 BTBL_MOUSE	P58544 mus musculu
42	9	2.2	497	1 WAS2_HUMAN	Q9y6w5 homo sapien
43	9	2.2	501	1 WASL_RAT	O08816 rattus norv
44	9	2.2	502	1 WASP_HUMAN	P42768 homo sapien
45	9	2.2	505	1 WASL_BOVIN	Q95107 bos taurus

ALIGNMENTS

RESULT 1				
UL33_RCMVM		STANDARD;	PRT;	387 AA.
ID	UL33_RCMVM			
AC	O12000;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	G-protein coupled receptor homolog R33.			
OS	Rat cytomagalovirus (strain Maastricht).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Betaherpesvirinae; Muromegalovirus.			
OX	NCBI_TaxID=79700;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98139136; PubMed=9499096;			
RA	Beisser P.S., Vink C., Van Dam J.G., Grauls G., Vanherle S.J.,			
RA	Bruggeman C.A.;			
RT	"The R33 G protein-coupled receptor gene of rat cytomagalovirus plays			
RT	an essential role in the pathogenesis of viral infection.";			
RL	J. Virol. 72:2352-2363(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20366325; PubMed=10906222;			
RA	Vink C., Beuken E., Bruggeman C.A.;			
RA	"Complete DNA sequence of the rat cytomagalovirus genome.";			
RL	J. Virol. 74:7656-7665(2000).			
CC	-I- FUNCTION: PLAYS AN IMPORTANT ROLE IN VIVO, IN PARTICULAR IN THE			
CC	DISSEMINATION TO OR REPLICATION IN THE SALIVARY GLAND.			
CC	-I- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-I- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; AF232689; AAC58815.1; -			
DR	GCRDB; GCR_1283; -			
DR	InterPro; IPR000276; GPCR_Rhodpsn.			
DR	Pfam; PF00001; 7tm_1; 1.			
DR	PROSITE; PS00237; G-PROTEIN_RECEP_FL_1; 1.			
DR	PROSITE; PS50262; G-PROTEIN_RECEP_FL_2; 1.			
DR	G-protein coupled receptor; Transmembrane.			
KW	DOMAIN 1 33			
FT	TRANSMEM 34 61			
FT	DOMAIN 62 71			
FT	TRANSMEM 72 94			
FT	DOMAIN 95 107			
FT	TRANSMEM 108 129			
FT	DOMAIN 130 150			
FT	TRANSMEM 151 169			
FT	DOMAIN 170 204			
FT	TRANSMEM 205 224			
FT	EXTRACELLULAR (POTENTIAL).			
FT	1 (POTENTIAL).			
FT	CYTOPLASMIC (POTENTIAL).			
FT	2 (POTENTIAL).			
FT	EXTRACELLULAR (POTENTIAL).			
FT	3 (POTENTIAL).			
FT	CYTOPLASMIC (POTENTIAL).			
FT	4 (POTENTIAL).			
FT	EXTRACELLULAR (POTENTIAL).			
FT	5 (POTENTIAL).			

```
FT DOMAIN 225 244 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 245 268 6 (POTENTIAL).
FT DOMAIN 269 285 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 286 309 7 (POTENTIAL).
FT DOMAIN 310 387 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 375 385 POLY-PRO.
FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 387 AA; 43159 MW; EF80D3F10344D6AE CRC64;

Query Match 2.5%; Score 10; DB 1; Length 387;
Best Local Similarity 100.0%; Pred. No. 0.17; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

QY 48 PPPPPPPPPN 57
Db 377 PPPPPPPPPN 386
|||||

RESULT 2
YPRO_OWEFU
ID YPRO_OWEFU STANDARD; PRT; 141 AA.
AC P21260; P21261;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical proline-rich protein (Fragment).
OS Owenia fusiformis.
OC Eukaryota; Metazoa; Annelida; Polychaeta; Palpata; Canaliculipata;
OC Sabellida; Oweniidae; Owenia.
OX NCBI_TaxID=6347;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=90147742; PubMed=2105723;
RA Bakalaria N., Collet J., Planells R., Thouveny Y., Fontes M.;
RT "Presence in invertebrate genomes of sequences characterized by the
repetition of the triplet CcPurine.";
RL Biochem. Biophys. Res. Commun. 166:66-73(1990).
DR PIR; A34043; A34043.
DR PIR; B34043; B34043.
KW Hypothetical protein; DNA-binding.
FT NON_TER 1 1
FT DOMAIN 9 58 POLY-PRO.
FT DNA_BIND 98 116 H-T-H MOTIF (POTENTIAL).
FT NON_TER 141 141
SQ SEQUENCE 141 AA; 15745 MW; B294E884D152BDB9 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 141;
Best Local Similarity 100.0%; Pred. No. 0.58;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56
Db 9 PPPPPPPPP 17
|||||

RESULT 3
Y084_RICPR
ID Y084_RICPR STANDARD; PRT; 165 AA.
AC Q92E63;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein RP084.
GN RP084.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
SEQUENCE FROM N.A.
RP SEQUENCE MADRID E;
RC STRAIN=MADRID E;
```

```

RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sicheritz-Ponten T., Almark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
mitochondria.";
RL Nature 396:133-140(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ235270; CAA14554.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 165 AA; 18232 MW; E844F1DF89C0A763 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56
Db 36 PPPPPPPPP 44
|||||

RESULT 4
TONB_XANCP
ID TONB_XANCP STANDARD; PRT; 223 AA.
AC Q34261.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TonB protein.
GN TONB.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=98037510; PubMed=9371459;
RA Wiggerich H.G., Klauke B., Koepf R., Priefer U.B., Puehler A.;
RT "Unusual structure of the tonB-exb DNA region of Xanthomonas
campestris pv. campestris: tonB, exbB, and exbD1 are essential for
ferric iron uptake, but exbD2 is not.";
RL J. Bacteriol. 179:7103-7110(1997).
CC -!- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER. FORMS A COMPLEX WITH THE ACCESSORY PROTEINS
CC EXB AND EXB2 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

DR EMBL; Z95386; CAB08610.1; --  
 DR InterPro; IPR003538; TonB.  
 DR PRINTS; PR01374; TonBPROTEIN.  
 KW Transport; Protein transport; Inner membrane; Periplasmic;  
 FT Transmembrane; Signal-anchor; Repeat.  
 FT DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 24 44 SIGNAL-ANCHOR (POTENTIAL).  
 FT DOMAIN 45 223 PERIPLASMIC (POTENTIAL).  
 FT DOMAIN 66 78 POLY-PRO.  
 FT DOMAIN 82 85 POLY-PRO.  
 FT DOMAIN 94 101 POLY-PRO.  
 SQ SEQUENCE 223 AA; 23598 MW; D7C6886FDF06FDA9 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 0.86; Indels 0; Gaps 0;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 P P P P P P P P P P 56  
 |||||  
 Db 66 P P P P P P P P P P 74

RESULT 5  
 VE4\_HPV08  
 ID VE4\_HPV08 STANDARD; PRT; 229 AA.  
 AC P06425;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-AUG-1992 (Rel. 23, Last annotation update)  
 DE Probable E4 protein.  
 OS Human papillomavirus type 8.  
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
 OC Papillomavirus.  
 OX NCBI\_TaxID=10579;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86200410; PubMed=3009874;  
 RA Fuchs P.G., Iftner T., Weninger J., Pfister H.;  
 RT "Epidermodyplasia verruciformis-associated human papillomavirus 8:  
 RL genomic sequence and comparative analysis.";  
 RL J. Virol. 58:626-634(1986).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M12737; --; NOT\_ANNOTATED\_CDS.  
 DR PIR; A03677; W4WL8.  
 KW Early protein.  
 SQ SEQUENCE 229 AA; 24205 MW; 5A5366640FA845C2 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 229;  
 Best Local Similarity 100.0%; Pred. No. 0.88;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 P P P P P P P P P P 57  
 |||||  
 Db 178 P P P P P P P P P P 186

RESULT 6  
 HXB4\_MOUSE  
 ID HXB4\_MOUSE STANDARD; PRT; 250 AA.  
 AC P10284;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Homeobox protein Hox-B4 (Hox-2.6).  
 GN HOXB4 OR HOXB-4 OR HOX-2.6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89091992; PubMed=2463210;  
 RA Graham A., Papalopulu N., Lorimer J., Mcvey J.H., Tuddenham E.G.D.,  
 RA Krumlauf R.;  
 RT "Characterization of a murine homeo box gene, Hox-2.6, related to the  
 RL Drosophila Deformed gene.";  
 RL Genes Dev. 2:1424-1438(1988).  
 CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.  
 CC "DEFORMED" SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M36654; AAA37848.1; --  
 DR PIR; A31757; A31757.  
 DR HSP; P02833; 9ANT.  
 DR TRANSFAC; T01726; --  
 DR MGD; MGI:96185; Hoxb4.  
 DR InterPro; IPR001827; Antennapedia.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 1.  
 DR PRINTS; PR00025; ANTENNAPEDIA.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
 DR PROSITE; PS50071; HOMEBOX\_2; 1.  
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 FT DOMAIN 15 138 PRO-RICH; PART OF THE TRANSCRIPTIONAL  
 FT DOMAIN 71 86 ACTIVATION DOMAIN.  
 FT DOMAIN 140 145 POLY-PRO.  
 FT DNA\_BIND 161 220 ANTP-TYPE HEXAPEPTIDE.  
 FT SEQUENCE 250 AA; 27519 MW; D09D477A0E585BE6 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 0.94;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 P P P P P P P P P P 56  
 |||||  
 Db 73 P P P P P P P P P P 81

RESULT 7  
 HXB4\_HUMAN  
 ID HXB4\_HUMAN STANDARD; PRT; 251 AA.  
 AC P17483; O9NTA0;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein Hox-B4 (Hox-2F) (Hox-2.6).  
 GN HOXB4 OR HOX2F.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91153613; PubMed=1981366;  
 RA Peverali F.A., D'Esposito M., Acampora D., Bunone G., Negri M.,  
 RA Faiella A., Stornaiuolo A., Pannese M., Migliacchio E., Simeone A.,  
 RA Valle G.D., Boncinelli E.;  
 RT "Expression of HOX homeogenes in human neuroblastoma cell culture  
 lines.";  
 RL Differentiation 45:61-69(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX Kidd K.K., Busyina V., Demille M.M.C., Speed W.C., Ruggeri V.,  
 RA Kidd J.R., Pakstis A.J.;  
 RT "Overall linkage disequilibrium in 33 populations for highly  
 RT informative multistate haplotypes spanning the HOXB gene cluster.";  
 RL Am. J. Hum. Genet. 67:235-235(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20538492; PubMed=11085749;  
 RA Giannola D.M., Shlomchik W.D., Jegathesan M., Liebowitz D.,  
 RA Abrams C.S., Kadesch T., Dancis A., Emerson S.G.;  
 RT "Hematopoietic expression of HOXB4 is regulated in normal and leukemic  
 RT stem cells through transcriptional activation of the HOXB4 promoter by  
 RT upstream stimulating factor (USF)-1 and USF-2.";  
 RL J. Exp. Med. 192:1479-1490(2000).  
 RN [4]  
 RP SEQUENCE OF 6-251 FROM N.A.  
 RX TISSUE=Testis;  
 RA Bloecker H., Boecher M., Brandt P., Mewes H.-W., Gassenhuber J.,  
 RA Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 160-227 FROM N.A.  
 RX TISSUE=Placenta;  
 RA MEDLINE=89378558; PubMed=2570724;  
 RA Giampaolo A., Acampora D., Zappavigna V., Pannese M.,  
 RA D'Esposito M., Care A., Faiella A., Stornaiuolo A., Russo G.,  
 RA Simeone A., Boncinelli E., Peschle C.;  
 RT "Differential expression of human HOX-2 genes along the anterior-  
 RT posterior axis in embryonic central nervous system.";  
 RL Differentiation 40:191-197(1989).  
 RN [6]  
 RP SEQUENCE OF 162-227 FROM N.A.  
 RX MEDLINE=90215256; PubMed=2576652;  
 RA Boncinelli E., Acampora D., Pannese M., D'Esposito M., Somma R.,  
 RA Gaudino G., Stornaiuolo A., Caffero M., Faiella A., Simeone A.;  
 RT "Organization of human class I homeobox genes.";  
 RL Genome 31:745-756(1989).  
 CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
 CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
 CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE EMBRYOS AND FETUSES AT  
 CC 5-9 WEEKS FROM CONCEPTION.  
 CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.  
 CC "DEFORMED" SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X16174; CAA34296.1; -  
 DR EMBL; AF287967; AAG31554.1; -  
 DR EMBL; AF307160; AAG45052.1; -  
 DR EMBL; AL137449; CAB70742.1; -  
 DR PIR; B60492; B60492.  
 DR PIR; C37042; C37042.

DR PIR; S15543; S15543.  
 DR HSP; P02833; 9ANT.  
 DR TRANSFAC; T01727; -.  
 DR MIM; 142965; -.  
 DR InterPro; IPR001827; Antennapedia.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 1.  
 DR PRINTS; PR00025; ANTENNAPEIDIA.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS00032; ANTENNAPEIDIA; 1.  
 DR PROSITE; PS00071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 FT DOMAIN 15 139 PRO-RICH; PART OF THE TRANSCRIPTIONAL  
 FT DOMAIN 71 87 POLY-PRO.  
 FT DOMAIN 141 146 ANTP-TYPE HEXAPEPTIDE.  
 FT DNA\_BIND 162 221 HOMEBOX.  
 SQ SEQUENCE 251 AA; 27604 MW; 23EE1D0DDCCE2DB4 CRC64;  
 QY 48 P P P P P P P P P P 56  
 DB 73 P P P P P P P P P P 81  
 Query Match 2.2%; Score 9; DB 1; Length 251;  
 Best Local Similarity 100.0%; Pred. No. 0.95;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 8  
 TRPC\_MYCTU  
 ID TRPC\_MYCTU STANDARD; PRT; 272 AA.  
 AC 06129;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Indole-3-glycerol phosphate synthase (EC 4.1.1.48) (IGPS).  
 GN TRPC OR RV1611 OR MT1646 OR MTCV01B2.03.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Mycobacterium.  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,  
 RA Rulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Esmolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: 1-(2-carboxyphenylamino)-1-deoxy-D-ribose 5-  
 CC phosphate = 1-(indol-3-yl)glycerol 3-phosphate + CO(2) + H(2)O.

CC -|- PATHWAY: FOURTH STEP IN BIOSYNTHESIS OF TRYPTOPHAN.  
 CC -|- SIMILARITY: BELONGS TO THE TRPC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z95554; CAB08905.1; -  
 CC EMBL; AB007029; AAK45915.1; ALT\_INT.  
 CC HSSP; P00909; IPII.  
 CC TIGR; MT1646; -  
 CC Tuberculin; RV1611; -  
 CC InterPro; IPR003009; FMN\_enzyme.  
 CC InterPro; IPR001468; IGPS.  
 CC Pfam; PF00218; IGPS; 1.  
 CC ProDom; PD001511; IGPS; 1.  
 CC PROSITE; PS00614; IGPS; 1.  
 CC TrpTophan biosynthesis; Lyase; Decarboxylase; Complete proteome.  
 KW DOMAIN 32 37 POLY-ALA.  
 FT CONFLICT 164 164 T -> I (IN REF. 2).  
 SQ SEQUENCE 272 AA; 28023 MW; 9CA29D0F0FAC76C2 CRC64;

Query Match 2.2%; Score 9; DB 1; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 AAAAAAPPP 148  
 |  
 Db 32 AAAAAAPPP 40

RESULT 9  
 ID FASL\_HUMAN STANDARD; PRT; 281 AA.  
 AC P48023;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE FAS antigen ligand (Apoptosis antigen ligand) (APTL) (CD178 antigen).  
 GN TNFSF6 OR APTLIG1 OR FASL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95105731; PubMed=7528780;  
 RA Alderson M.;  
 RT "Fas ligand mediates activation-induced cell death in human T  
 RT lymphocytes.";  
 RL J. Exp. Med. 181:71-77(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95127560; PubMed=7826947;  
 RA Takahashi T., Tanaka M., Inazawa J., Abe T., Suda T., Nagata S.;  
 RT "Human Fas ligand: gene structure, chromosomal location and species  
 RT specificity.";  
 RL Int. Immunol. 6:1567-1574(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Schaeublein C.E., Poehlmann R., Philippson P., Eibel H.;  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95071350; PubMed=7980502;  
 RA Mita E., Hayashi N., Iio S., Takehara T., Hijioaka T., Kasahara A.,  
 RA Fusamoto H., Kamada T.;  
 RT "Role of Fas ligand in apoptosis Induced by hepatitis C virus

RT Infection.";  
 RL Biochem. Biophys. Res. Commun. 204:468-474(1994).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Wilkinson J.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RC TISSUE=Blood;  
 RA Matsumura M., Nakanishi Y., Ohba Y.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -|- FUNCTION: CYTOKINE THAT BINDS TO FAS ANTIGEN, A RECEPTOR THAT  
 CC TRANSDUCES THE APOPTOTIC SIGNAL INTO CELLS. MAY BE INVOLVED IN  
 CC CYTOTOXIC T CELL MEDIATED APOPTOSIS AND IN T CELL DEVELOPMENT.  
 CC FAS-ANTIGEN MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE INDUCTION OF  
 CC PERIPHERAL TOLERANCE, IN THE ANTIGEN-STIMULATED SUICIDE OF MATURE  
 CC T CELLS, OR BOTH.  
 CC -|- SUBUNIT: HOMOTRIMER (PROBABLE).  
 CC -|- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MAY BE RELEASED  
 CC INTO THE EXTRACELLULAR FLUID, PROBABLY BY CLEAVAGE FORM THE CELL  
 CC SURFACE.  
 CC -|- SIMILARITY: BELONGS TO THE TUMOR NECROSIS FACTOR FAMILY.  
 CC -|- DATABASE: NAME=PRO; NOTE=CD guide CD178 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/guide/338769674\_g.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X89102; CAA61474.1; -  
 CC EMBL; U08137; AAC50071.1; -  
 CC EMBL; U11821; AAC50124.1; -  
 CC EMBL; D38122; BAA07320.1; -  
 CC EMBL; Z96050; CAB09424.1; -  
 CC EMBL; AB013303; BAA32542.1; -  
 CC HSSP; P06804; 2TNF.  
 CC MTM; 134638; -  
 CC InterPro; IPR003263; TNF\_5.  
 CC InterPro; IPR003636; TNF\_abcd.  
 CC InterPro; IPR000478; TNF\_family.  
 CC Pfam; PF00229; TNF; 1.  
 CC ProDom; PD002012; TNF\_abcd; 1.  
 CC ProDom; PD008600; TNF\_5; 1.  
 CC SMART; SM00207; TNF; 1.  
 CC PROSITE; PS00251; TNF\_1; 1.  
 CC PROSITE; PS00049; TNF\_2; 1.  
 CC Cytokine; Transmembrane; Glycoprotein; Signal-anchor; Apoptosis;  
 KW Antigen.  
 FT DOMAIN 1 80 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 81 102 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 FT DOMAIN 103 281 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 4 70 PRO-RICH.  
 FT DOMAIN 45 65 POLY-PRO.  
 FT DISULFID 202 233 BY SIMILARITY.  
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 281 AA; 31485 MW; A8A6EB358246E9BB CRC64;

Query Match 2.2%; Score 9; DB 1; Length 281;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 pppppppppp 56  
 |  
 Db 45 pppppppppp 53



FT DOMAIN 191 197 POLY-PRO.  
 FT DOMAIN 213 222 POLY-PRO.  
 FT DOMAIN 239 243 POLY-PRO.  
 FT CONFLICT 8 MISSING (IN REF. 2).  
 FT CONFLICT 65 K -> E (IN REF. 2).  
 FT CONFLICT 211 S -> N (IN REF. 2).  
 SQ SEQUENCE 289 AA; 31193 MW; A8236F80791CE52B CRC64;

Query Match 2.2%; Score 9; DB 1; Length 289;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
 |||||

Db 213 PPPPPPPPP 221

RESULT 12  
 HXD8\_HUMAN STANDARD; PRT; 290 AA.  
 AC P13378;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein Hox-D8 (Hox-4E) (Hox-5.4).  
 GN HOXD8 OR HOX4E.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Birren B., Linton L., Nusbaum C., Lander E.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Kosaki K., Kosaki R., Suzuki T., Yoshihashi H., Sasaki K., Matsuo N.;  
 RT "A complete mutation analysis panel of human HOX genes."  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE OF 193-287 FROM N.A.  
 RX MEDLINE=89306602; PubMed=2568311;  
 RA Oliver G., Sidell N., Fiske N., Heinzmann C., Mohandas T.,  
 RA Sparkes R.S., de Robertis E.M.;  
 RT "Complementary homeo protein gradients in developing limb buds."  
 RL Genes Dev. 3:641-650(1989).

CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
 A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH  
 SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AC009336; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; AY014304; AAG42152.1; -  
 DR EMBL; AY014303; AAG42152.1; JOINED.  
 DR EMBL; X15507; CAA33529.1; -  
 DR PIR; B32830; B32830.  
 DR PIR; S05957; S05957.  
 DR HSP; P02833; 9ANT.  
 DR TRANSEC; T03332; -  
 DR MIM; 142985; -  
 DR InterPro; IPR001827; Antennapedia.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; homeobox; 2.

DR PRINTS; PR00025; ANTENNAPEDIA.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1. PARTIAL.  
 DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
 DR PROSITE; PS0071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 FT DOMAIN 15 23 POLY-ALA.  
 FT DOMAIN 45 50 POLY-ALA.  
 FT DOMAIN 109 123 POLY-PRO.  
 FT DNA\_BIND 197 256 HOMEBOX.  
 FT CONFLICT 287 287 G -> A (IN REF. 3).  
 SQ SEQUENCE 290 AA; 31910 MW; 75FF95A73E2AA85F CRC64;

Query Match 2.2%; Score 9; DB 1; Length 290;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
 |||||

Db 115 PPPPPPPPP 123

RESULT 13  
 SMN\_HUMAN STANDARD; PRT; 294 AA.  
 ID SMN\_HUMAN  
 AC Q16637; Q13119;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Survival motor neuron protein (Component of gems 1) (Gemin1).  
 GN (SMN1 OR SMN OR SMNT) AND (SMN2 OR SMNC).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANT SMA CYS-272.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=95112343; PubMed=7813012;  
 RA Lefebvre S., Buerklen L., Reboullet S., Clermont O., Bulet P.,  
 RA Viollet L., Benichou B., Cruaud C., Millasseau P., Zeviani M.,  
 RA le Paslier D., Frezal J., Cohen D., Weissenbach J., Munnich A.,  
 RA Melki J.;  
 RT "Identification and characterization of a spinal muscular atrophy-  
 RT determining gene."  
 RL Cell 80:155-165(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96435930; PubMed=8838816;  
 RA Buerklen L., Lefebvre S., Clermont O., Bulet P., Viollet L.,  
 RA Cruaud C., Munnich A., Melki J.;  
 RT "Structure and organization of the human survival motor neurone (SMN)  
 RT gene."  
 RL Genomics 32:479-482(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Chen Q., Baird S.D., Besner-Johnston A., Farahani R., Xuan J.Y.,  
 RA Kang X., Lefebvre C., Korneluk R.G., Mackenzie A.E.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Jones K., Graves T., McPherson J.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=95367018; PubMed=7639755;  
 RA Gennarelli M., Lucarelli M., Capon F., Pizzuti A., Merlini L.,  
 RA Angelini C., Novelli G., Dallapiccola B.;  
 RT "Survival motor neuron gene transcript analysis in muscles from spinal

RT muscular atrophy patients.";  
RL Biochem. Biophys. Res. Commun. 213:342-348(1995).  
RN [6]  
RX SEQUENCE OF 2-294 FROM N.A.  
RP MEDLINE=96018066; PubMed=7552146;  
RA van der Steege G., Draaijers T.G., Grootsholten P.M., Osinga J.,  
RA Anzevino R., Velona I., Den Dunnen J.T., Scheffer H., Brahe C.,  
RA van Ommen G.J.;  
RT "A provisional transcript map of the spinal muscular atrophy (SMA)  
RT critical region.";  
RL Eur. J. Hum. Genet. 3:87-95(1995).  
RN [7]  
RP SUBCELLULAR LOCATION  
RX MEDLINE=96312953; PubMed=8670859;  
RA Liu Q., Dreyfuss G.;  
RT "A novel nuclear structure containing the survival of motor neurons  
RT protein.";  
RL EMBO J. 15:3555-3565(1996).  
RN [8]  
RP FUNCTION.  
RX MEDLINE=99059497; PubMed=9845364;  
RA Pellizzoni L., Kataoka N., Charroux B., Dreyfuss G.;  
RT "A novel function for SMN, the spinal muscular atrophy disease gene  
RT product, in pre-mRNA splicing.";  
RL Cell 95:615-624(1998).  
RN [9]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=97402201; PubMed=9259265;  
RA Coovert D.D., Le T.T., McAndrew P.E., Strasswimmer J., Crawford T.O.,  
RA Mendell J.R., Coulson S.E., Androphy E.J., Prior T.W., Burghes A.H.;  
RT "The survival motor neuron protein in spinal muscular atrophy.";  
RL Hum. Mol. Genet. 6:1205-1214(1997).  
RN [10]  
RP INTERACTION WITH SIP1.  
RX MEDLINE=97462902; PubMed=9323129;  
RA Liu Q., Fischer U., Wang F., Dreyfuss G.;  
RT "The spinal muscular atrophy disease gene product, SMN, and its  
RT associated protein SIP1 are in a complex with spliceosomal snRNP  
RT proteins.";  
RL Cell 90:1013-1021(1997).  
RN [11]  
RP STRUCTURE BY NMR OF 82-169.  
RX MEDLINE=20577270; PubMed=11135666;  
RA Selenko P., Sprangers R., Stier G., Buhler D., Fischer U., Sattler M.;  
RT "SMN tudor domain structure and its interaction with the Sm  
RT proteins.";  
RL Nat. Struct. Biol. 8:27-31(2001).  
RN [12]  
RP VARIANT SMA VAL-279.  
RX MEDLINE=97227297; PubMed=9147655;  
RA Talbot K., Ponting C.P., Theodosiou A.M., Rodrigues N.R., Surtees R.,  
RA Mountford R., Davies K.E.;  
RT "Missense mutation clustering in the survival motor neuron gene: a  
RT role for a conserved tyrosine and glycine rich region of the protein  
RT in RNA metabolism?";  
RL Hum. Mol. Genet. 6:497-500(1997).  
RN [13]  
RP VARIANTS SMA ILE-262 AND ILE-274.  
RX MEDLINE=97301781; PubMed=9158159;  
RA Hahnen E., Schoenling J., Rudnik-Schoeneborn S., Raschke H.,  
RA Zerres K., Wirth B.;  
RT "Missense mutations in exon 6 of the survival motor neuron gene in  
RT patients with spinal muscular atrophy (SMA).";  
RL Hum. Mol. Genet. 6:821-825(1997).  
RN [14]  
RP VARIANTS SMA LEU-245 AND CYS-272.  
RX MEDLINE=20194955; PubMed=10732817;  
RA Rochette C.F., Surh L.C., Ray P.N., McAndrew P.E., Prior T.W.,  
RA Burghes A.H.M., Vanasse M., Smard L.R.;  
RT "Molecular diagnosis of non-deletion SMA patients using quantitative  
RT PCR of SMN exon 7.";  
RL Neurogenetics 1:141-147(1997).  
RN [15]  
RP VARIANT SMA GLY-2.  
RX MEDLINE=99057509; PubMed=9837824;  
RA Parsons D.W., McAndrew P.E., Iannaccone S.T., Mendell J.R.,  
RA Burghes A.H., Prior T.W.;  
RT "Intragenic telSMN mutations: frequency, distribution, evidence of a  
RT founder effect, and modification of the spinal muscular atrophy  
RT phenotype by cenSMN copy number.";  
RL Am. J. Hum. Genet. 63:1712-1723(1998).  
RN [16]  
RP VARIANT SMA CYS-279.  
RX MEDLINE=20194940; PubMed=10732802;  
RA Wang C.H., Papendick B.D., Bruinsma P., Day J.K.;  
RT "Identification of a novel missense mutation of the smnt gene in two  
RT siblings with spinal muscular atrophy.";  
RL Neurogenetics 1:273-276(1998).  
CC -!- FUNCTION: THE SMN COMPLEX PLAYS AN ESSENTIAL ROLE IN SPLICEOSOMAL  
CC SNRP ASSEMBLY IN THE CYTOPLASM, AND IS REQUIRED FOR PRE-MRNA  
CC SPLICING IN THE NUCLEUS.  
CC -!- SUBUNIT: FORMS A STABLE HETEROMERIC COMPLEX WITH SURVIVAL OF MOTOR  
CC NEURON PROTEIN INTERACTING PROTEIN 1 (SIP1/GEMIN2), GEMIN3 AND  
CC GEMIN4. THIS COMPLEX IS THOUGHT TO PLAY AN ESSENTIAL ROLE IN  
CC BIOGENESIS OF THE SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES  
CC (SNRNPs). INTERACTS WITH FIBRILLARIN AND WITH SEVERAL SPLICEOSOMAL  
CC SNRP CORE SM PROTEINS, INCLUDING B/B', SM D AND SM E PROTEINS AND  
CC WITH ILF3.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR. LOCALIZED IN  
CC SUBNUCLEAR STRUCTURES NEXT TO COILED BODIES, CALLED GEMS, WHICH  
CC ARE HIGHLY ENRICHED IN SPLICEOSOMAL SNRNPs.  
CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS: SMN/FULL-SMN (SHOWN HERE), SMN-  
CC DELTA5/ISO5-SMN, SMN-DELTA7/ISO7-SMN AND SMN-DELTA57/ISO57-SMN;  
CC SEEM TO BE PRODUCED BY ALTERNATIVE SPLICING. THE SMN-DELTA7  
CC ISOFORM IS THOUGHT TO BE A NONFUNCTIONAL PROTEIN THAT LACKS THE  
CC CAPACITY TO OLIGOMERIZE AND, THUS CANNOT INTERACT WITH SM  
CC PROTEINS. SMN1 PRIMARILY PRODUCES SMN WHEREAS SMN2 EXPRESSES  
CC ABUNDANT LEVELS OF SMN-DELTA7.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES.  
CC EXPRESSED AT HIGH LEVELS IN BRAIN, KIDNEY AND LIVER, MODERATE  
CC LEVELS IN SKELETAL AND CARDIAC MUSCLE, AND LOW LEVELS IN  
CC FIBROBLASTS AND LYMPHOCYTES. ALSO SEEN AT HIGH LEVELS IN SPINAL  
CC CORD.  
CC -!- DISEASE: MUTATED FORMS OF SMN HAVE BEEN FOUND IN INDIVIDUALS WITH  
CC SPINAL MUSCULAR ATROPHIES (SMA). SMAS ARE FATAL AUTOSOMAL  
CC RECESSIVE DISORDERS SUBCLASSIFIED AS TYPE 1 (WERNIG-HOFFMANN  
CC DISEASE), TYPE II (INTERMEDIATE FORM), AND TYPE III (HOEHLFART-  
CC KUGELBERG-WELANDER DISEASE) BASED UPON THE AGE OF ONSET AND  
CC CLINICAL SEVERITY. THESE NEURODEGENERATIVE DISORDERS ARE  
CC CHARACTERIZED BY DEGENERATION OF LOWER MOTOR NEURONS, LEADING TO  
CC PROGRESSIVE PARALYSIS MUSCULAR ATROPHY. CONCERNS 1 IN 6000  
CC NEWBORNS  
CC -!- MISCELLANEOUS: THE SMN GENE IS PRESENT IN TWO HIGHLY HOMOLOGOUS  
CC AND FUNCTIONAL COPIES (TEL-SMN AND CEN-SMN). THE TELOMERIC COPY OF  
CC SMN GENE SEEMS TO BE THE SMA-DETERMINING GENE WHILE THE  
CC CENTROMERIC COPY SEEMS UNAFFECTED.  
CC -!- SIMILARITY: BELONGS TO THE SMN FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 TUDOR DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U43883; AAC50473.1; -;  
DR EMBL; U43876; AAC50473.1; JOINED.  
DR EMBL; U43877; AAC50473.1; JOINED.  
DR EMBL; U43878; AAC50473.1; JOINED.  
DR EMBL; U43880; AAC50473.1; JOINED.  
DR EMBL; U43881; AAC50473.1; JOINED.  
DR EMBL; U43882; AAC50473.1; JOINED.  
DR EMBL; U18423; AAC66242.1; -;  
DR EMBL; U80017; AAC52048.1; -;



DR EMBL; AC005031; AAC62262.1; -;  
DR EMBL; U21914; AAA64505.1; -;  
DR MIM; 600354; -;  
DR MIM; 601627; -;  
DR MIM; 253300; -;  
DR MIM; 253400; -;  
DR MIM; 253550; -;  
DR PDB; 1G5V; 02-MAY-01.

Query Match 2.2%; Score 9; DB 1; Length 294;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56  
| | | | | | | | | |  
DB 217 PPPPPPPPP 225

RESULT 14  
HX1M\_CHICK STANDARD; PRT; 297 AA.  
AC O93367;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Homeobox protein Hox-11L2 (Homeobox TLX-3).  
GN HOX11L2 OR TLX3.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN-RHODE ISLAND RED;  
RA Logan C.C., Wingate R.J.T., McKay I.J., Lumsden A.;  
RT "Hlx-1 and Hlx-3 homeobox gene expression in cranial sensory ganglia  
RT and hindbrain of the chick embryo: markers of patterned  
RT connectivity".  
RL J. Neurosci. 18:5389-5402(1998).  
CC -1- FUNCTION: SEEMS TO BE INVOLVED IN THE DEVELOPMENT OF CRANIAL  
CC SENSORY INNERVATION FROM PERIPHERAL GANGLIA.  
CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).  
CC -1- TISSUE SPECIFICITY: EXPRESSION IS RESTRICTED TO NEURONES IN THE  
CC PERIPHERAL AND CENTRAL NERVOUS SYSTEM.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; AF071875; AAC23901.1; -;  
DR HSSP; P14653; 1b72.  
DR InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00025; ANTENNAPEDIA.  
DR PRINTS; PR00024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
KW Homeobox; DNA-binding; Homeobox protein; Developmental protein.  
FT DOMAIN 33 50 POLY-PRO.  
FT DOMAIN 147 150 POLY-ALA.  
FT DNA\_BIND 172 231 HOMEBOX.  
SQ SEQUENCE 297 AA; 32369 MW; 0C7C9A73E371A38E CRC64;

Query Match 2.2%; Score 9; DB 1; Length 297;

Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 PPPPPPPPP 56  
| | | | | | | | | |  
DB 33 PPPPPPPPP 41

RESULT 15  
HX4A\_CHICK STANDARD; PRT; 309 AA.  
ID HX4A\_CHICK  
AC P17277;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-JUL-1993 (Rel. 26, Last annotation update)  
DE Homeobox protein Hox-A4 (Chox-1.4).  
GN HOXA4 OR CHOX-1.4.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90245562; PubMed=1970866;  
RA Sasaki H., Yokoyama E., Kuroiwa A.;  
RT "Specific DNA binding of the two chicken Deformed family homeodomain  
RT proteins, Chox-1.4 and Chox-a".  
RL Nucleic Acids Res. 18:1739-1747(1990).  
RN [2]  
RP SEQUENCE OF 207-273 FROM N.A.  
RX STRAIN-COMET HUBBARD HYBRID;  
RA Scotting P.J., Hewitt M., Keynes R.J.;  
RT "Isolation and analysis of chick homeobox cDNA clones".  
RL Nucleic Acids Res. 18:3999-3999(1990).  
CC -1- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF  
CC A DEVELOPMENTAL REGULATORY SYSTEM ON THE ANTERIOR-POSTERIOR AXIS.  
CC A SPECIFIC POSITIONAL IDENTITIES ON THE 5'-FLANKING SEQUENCE OF ITS CODING  
CC REGION WITH VARIOUS AFFINITIES. THE CONSENSUS SEQUENCES OF THE  
CC HIGH AND LOW AFFINITY BINDING SITES ARE TAATGA(C/G) AND CTAATTTT.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- MISCELLANEOUS: THE PROLINE STRETCH WORKS AS A PART OF THE  
CC TRANSCRIPTIONAL ACTIVATION DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.  
CC "DEFORMED" SUBFAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL; X52670; CAA36896.1; -;  
DR EMBL; X52747; CAB57949.1; -;  
DR PIR; S09257; S09257.  
DR PIR; S10883; S10883.  
DR HSSP; P02833; 9ANT.  
DR TRANSFAC; T00128; -;  
DR InterPro; IPR001827; Antennapedia.  
DR InterPro; IPR001356; Homeobox.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00025; ANTENNAPEDIA.  
DR PRINTS; PR00024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
DR PROSITE; PS00032; ANTENNAPEDIA; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 15:12:44 ; Search time 31.84 Seconds  
(without alignments)  
2178.736 Million cell updates/sec

Title: US-09-631-863A-2  
Perfect score: 401  
Sequence: 1 MRNRKVLTKRKRSRGQD.....DNCPSAKSKSSPAGNSPAPL 401

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 562222 seqs, 172994929 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries

Database : SPTREMBL19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhcc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	342	85.3	342	4	Q9UPV1 homo sapien
2	325	81.0	325	4	Q96A68
3	15	3.7	231	11	Q9EQ11
4	10	2.5	212	10	Q9AYL5
5	10	2.5	240	10	Q9XFB1
6	10	2.5	250	5	Q20327
7	10	2.5	268	13	Q9DGN3
8	10	2.5	276	10	Q9SNS1
9	10	2.5	384	11	Q9CUF6
10	10	2.5	409	10	Q9SBN1
11	10	2.5	494	10	Q9FXA1
12	10	2.5	1018	5	O61701
13	10	2.5	1432	5	Q95ZV4
14	9	2.2	33	6	O02832
15	9	2.2	48	5	Q9VWS6
16	9	2.2	61	11	O08549

17	9	2.2	67	11	Q9ZLI8	Q9ZLI8 rattus norv
18	9	2.2	83	5	Q9GTI9	Q9GTI9 heterodera
19	9	2.2	90	12	O41048	O41048 paramelicu
20	9	2.2	96	5	Q9VTD0	Q9VTD0 drosophila
21	9	2.2	109	5	Q9BLB3	Q9BLB3 caenorhabdi
22	9	2.2	124	10	Q9FN57	Q9FN57 arabidopsis
23	9	2.2	126	12	O11299	O11299 molluscum c
24	9	2.2	127	4	Q9BZP9	Q9BZP9 mus sapien
25	9	2.2	135	10	Q9CAH6	Q9CAH6 arabidopsis
26	9	2.2	142	5	Q00485	Q00485 hydra sp. m
27	9	2.2	147	11	Q9D2F0	Q9D2F0 mus musculu
28	9	2.2	148	11	O64371	O64371 rattus norv
29	9	2.2	149	5	Q00484	Q00484 hydra sp. m
30	9	2.2	158	12	Q9DVM0	Q9DVM0 pluteia xy
31	9	2.2	161	10	Q9M0N0	Q9M0N0 arabidopsis
32	9	2.2	162	5	Q23137	Q23137 caenorhabdi
33	9	2.2	163	6	Q95MP5	Q95MP5 bos taurus
34	9	2.2	163	11	Q9CU08	Q9CU08 mus musculu
35	9	2.2	169	11	O35315	O35315 mus musculu
36	9	2.2	172	5	Q00487	Q00487 hydra sp. m
37	9	2.2	172	5	Q23248	Q23248 caenorhabdi
38	9	2.2	173	11	Q9D487	Q9D487 mus musculu
39	9	2.2	176	5	Q16990	Q16990 acropora do
40	9	2.2	179	11	Q9D0Y6	Q9D0Y6 mus musculu
41	9	2.2	182	12	Q98324	Q98324 molluscum c
42	9	2.2	183	4	Q9UQL5	Q9UQL5 mus sapien
43	9	2.2	186	5	Q00486	Q00486 hydra sp. m
44	9	2.2	192	4	Q9UHL1	Q9UHL1 mus sapien
45	9	2.2	195	10	Q43460	Q43460 glycine max

ALIGNMENTS

RESULT 1

Q9UPV1	ID	Q9UPV1	PRELIMINARY;	PRT;	342 AA.
AC	Q9UPV1;				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DE	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	K1AA1051	PROTEIN (FRAGMENT).			
GN	K1AA1051.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=BRAIN;				
RX	MEDLINE=9397452; PubMed=10470851;				
RA	Kikuno R., Nagase T., Ishikawa K., Hirose M., Miyajima N.,				
RA	Tanaka A., Kotani H., Nomura N., Ohara O.;				
RT	"Prediction of the coding sequences of unidentified human genes. XIV.				
RT	The complete sequences of 100 new cDNA clones from brain which code				
RL	for large proteins in vitro.;"				
RL	DNA Res. 6:197-205(1999).				
DR	EMBL; AB028974; BAA83003.1; -				
DR	InterPro; IPR001878; Znf_CCHC.				
DR	Pfam; PF00098; zf-CCHC; 1.				
DR	SMART; SM00343; Znf_C2HC; 1.				
KW	zinc-finger.				
FT	NON_TER				
SQ	SEQUENCE	342 AA;	38769 MW;	DC14C265B92541EB	CRC64;

Query Match	85.3%;	Score 342;	DB 4;	Length 342;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 342;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
QY	60	NNNSKHTGHKSACVPNMTERRRDELSEINNLRKVMKQSENNNLQSOVKLTTENTT	119	
Db	1	NNNSKHTGHKSACVPNMTERRRDELSEINNLRKVMKQSENNNLQSOVKLTTENTT	60	

QY 120 LREQVEPTPEDEDDDDIELRGAAAAAAPPPIEEBCPEDLPKFDGNDPMLAPFMAQCQIF 179  
 |||||  
 DB 61 LREQVEPTPEDEDDDDIELRGAAAAAAPPPIEEBCPEDLPKFDGNDPMLAPFMAQCQIF 120  
 |||||  
 QY 180 MEKTRDFSVDRVRVCFVTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQR 239  
 |||||  
 DB 121 MEKTRDFSVDRVRVCFVTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQR 180  
 |||||  
 QY 240 REVAKRKIRLRROGMGSVIDYSNPFQMTAQDLWDNEPALIDQYHEGLSDHIEELSHLEV 299  
 |||||  
 DB 181 REVAKRKIRLRROGMGSVIDYSNPFQMTAQDLWDNEPALIDQYHEGLSDHIEELSHLEV 240  
 |||||  
 QY 300 AKSLSALIGQCIIHIERRLARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTQ 359  
 |||||  
 DB 241 AKSLSALIGQCIIHIERRLARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTQ 300  
 |||||  
 QY 360 EEKERRKLNLCYCGTGGHYADNCPAKAKSSPAGNSPAPL 401  
 |||||  
 DB 301 EEKERRKLNLCYCGTGGHYADNCPAKAKSSPAGNSPAPL 342  
 |||||

## RESULT 2

Q96A68 ID Q96A68 PRELIMINARY; PRT; 325 AA.  
 AC Q96A68;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE MEF3 LIKE 1 (PATERNALLY EXPRESSED GENE 10 ORF1).  
 GN MEF3L1 OR PEG10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Nakamura Y., Furukawa Y.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=TESTIS;  
 RA Satoh S., Furukawa Y.;  
 RL "Isolation of MEF3 like gene 1.";  
 RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21218929; PubMed=11318613;  
 RA Ono R., Kobayashi S., Wagatsuma H., Alsaka K., Kohda T.,  
 RA Kaneko-Ishino T., Ishino F.;  
 RT "A Retrotransposon-Derived Gene, PEG10, Is a Novel Imprinted Gene  
 RT Located on Human Chromosome 7q21.";  
 RL Genomics 73:232-237(2001).  
 DR EMBL; AB049150; BAB68387.1; -;  
 DR EMBL; AB049834; BAB43951.1; -;  
 SQ SEQUENCE 325 AA; 36965 MW; 118E4CAF97F2A76 CRC64;

Query Match 81.0%; Score 325; DB 4; Length 325;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 MTERRRDSEIENLNREKVMKQSEENNLSQVOKLTEENTTLREQVEPTPEDEDDDDIE 136  
 |||||  
 DB 1 MTERRRDSEIENLNREKVMKQSEENNLSQVOKLTEENTTLREQVEPTPEDEDDDDIE 60  
 |||||  
 QY 137 LRGAAAAAAPPPIEEBCPEDLPKFDGNDPMLAPFMAQCQIFMEKSTRDFSVDRVRVCF 196  
 |||||  
 DB 61 LRGAAAAAAPPPIEEBCPEDLPKFDGNDPMLAPFMAQCQIFMEKSTRDFSVDRVRVCF 120  
 |||||  
 QY 197 VTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQRREVAKRKIRLRROGMGS 256  
 |||||

DB 121 VTSMTGTGAARWASAKLERSHYLMHNYPAFMEMKMHVFEDPQRREVAKRKIRLRROGMGS 180  
 |||||  
 QY 257 VIDYSNPFQMTAQDLWDNEPALIDQYHEGLSDHIEELSHLEVAKSLSALIGQCIIHIERR 316  
 |||||  
 DB 181 VIDYSNPFQMTAQDLWDNEPALIDQYHEGLSDHIEELSHLEVAKSLSALIGQCIIHIERR 240  
 |||||  
 QY 317 LARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTQEEKERRKLNLCYCGT 376  
 |||||  
 DB 241 LARAAAAARKPRSPRALVLPVPHIASHHQVDPTPEVGGARMRLTQEEKERRKLNLCYCGT 300  
 |||||  
 QY 377 GGHYADNCPAKAKSSPAGNSPAPL 401  
 |||||  
 DB 301 GGHYADNCPAKAKSSPAGNSPAPL 325  
 |||||

## RESULT 3

Q9EQ11 ID Q9EQ11 PRELIMINARY; PRT; 231 AA.  
 AC Q9EQ11;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE MYELIN EXPRESSION FACTOR-3-LIKE PROTEIN (FRAGMENT).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX Wolff J.-N., Koerting C., Scharltl M.;  
 RT "Ty3/Gypsy retrotransposon fossils in mammalian genomes: did they  
 RT evolve into new cellular functions?";  
 RL Mol. Biol. Evol. 18:286-270(2001).  
 DR EMBL; AF302691; AAG39979.1; -;  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00098; zf-CCHC; 1.  
 DR SMART; SM00343; Znf\_C2HC; 1.  
 KW Zinc-finger.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 231 AA; 26171 MW; DEAB2A2B624F3974 CRC64;

Query Match 3.7%; Score 15; DB 11; Length 231;  
 Best Local Similarity 100.0%; Pred. No. 3e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 DYSNPFQMTAQDLWD 273  
 |||||  
 DB 93 DYSNPFQMTAQDLWD 107  
 |||||

## RESULT 4

Q9AYL5 ID Q9AYL5 PRELIMINARY; PRT; 212 AA.  
 AC Q9AYL5;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 23.1 KDA PROTEIN.  
 GN OSJNBA0003019.2.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE;  
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,  
 RA Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,  
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblum T.V.,

```

RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNBa0003019 genomic sequence.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC060755; AAK00445.1; -.
DR InterPro: IPR000916; Bel_v_1.
DR ProDom: PD000531; Bel_v_1.
KW Hypothetical protein.
SQ SEQUENCE 212 AA; 23084 MW; BBA1D03553465083 CRC64;

Query Match 2.5%; Score 10; DB 10; Length 212;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 AAAAAA PPPP 149
DB 197 AAAAAA PPPP 206

RESULT 5
Q9XFB1 PRELIMINARY; PRT; 240 AA.
AC Q9XFB1.
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DE YABBY3 (At4G00180/FGN15_22).
GN YABBY3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99387979; PubMed=10457020;
RA Siegfried K.R., Eshed Y., Baum S.F., Otsuga D., Drews G.N.,
RA Bowman J.L.;
RT "Members of the YABBY gene family specify abaxial cell fate in
RT Arabidopsis.";
RL Development 126:4117-4128(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Shinn P.,
RA Tracy S.E., Banh J., Bowser L., Carinci P., Chung M.K.,
RA Goldsmith A.D., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Pham P.K., Quach H.L.,
RA Sakano H., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,
RA Toriumi M., Yamada K., Yu G., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF136540; AAD33717.1; -.
DR EMBL: AY037186; AAK59771.1; -.
SQ SEQUENCE 240 AA; 26338 MW; 4D90DA549B54A9AC CRC64;

Query Match 2.5%; Score 10; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPPN 57
DB 84 PPPPPPPPPN 93

RESULT 6
Q20327 PRELIMINARY; PRT; 250 AA.
ID Q20327
AC Q20327
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

```

```

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHETICAL 27.7 KDA PROTEIN.
GN F42C5.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Du Z.;
RT "The sequence of C. elegans cosmid F42C5.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL EMBL: U40799; AAA81484.1; -.
DR EMBL: U40799; AAA81484.1; -.
DR InterPro: IPR001467; Molybdopterin.
DR PROSITE: PS00551; MOLYBDOPTERIN_PROK_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 250 AA; 27685 MW; 477A9EC12E3C3585 CRC64;

Query Match 2.5%; Score 10; DB 5; Length 250;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 CPPPPPPPPP 56
DB 102 CPPPPPPPPP 111

RESULT 7
Q9DGN3 PRELIMINARY; PRT; 268 AA.
ID Q9DGN3
AC Q9DGN3
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE ESPIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COCHLEA;
RX MEDLINE=20427267; PubMed=10975527;
RA Zheng L., Sekerkova G., Vranich K., Tilney L.G., Mugnaini E.,
RA Bartles J.R.;
RT "The Deaf Jerker Mouse Has a Mutation in the Gene Encoding the Espin
RT Actin-Bundling Proteins of Hair Cell Stereocilia and Lacks Espins.";
RL Cell 102:377-385(2000).
DR EMBL: AF239885; AAF98133.1; -.
DR InterPro: IPR000381; Inhibin_beta.
DR InterPro: IPR002965; P_rich_extensn.
DR InterPro: IPR000633; Vinculin_2.
DR InterPro: IPR003124; WH2.
DR Pfam: PF02205; WH2; 1.
DR PRINTS: PR00671; INHIBINBB.
DR PRINTS: PR01217; PRICHEXTNSN.

```

```

DR PRINTS; PRO0806; VINCULIN.
DR SMART; SM00246; WH2; 1.
FT NON_TER 1
SQ SEQUENCE 268 AA; 29103 MW; F815B4828B2BB9A2 CRC64;

Query Match 2.5%; Score 10; DB 13; Length 268;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 47 CPPPPPPPPPP 56
   |||||
Db 26 CPPPPPPPPPP 35

RESULT 8
Q9SNS1 ID Q9SNS1 PRELIMINARY; PRT; 276 AA.
AC Q9SNS1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
RT clone:P0535G04.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DE EMBL; AP000399; BAA83555.1; -.
KW Hypothetical protein.
SQ SEQUENCE 276 AA; 29463 MW; 62B4AA1069536447 CRC64;

Query Match 2.5%; Score 10; DB 10; Length 276;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 AAAAAAAPP 149
   |||||
Db 79 AAAAAAAPP 88

RESULT 9
Q9CUF6 ID Q9CUF6 PRELIMINARY; PRT; 384 AA.
AC Q9CUF6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 4930577N17RIK PROTEIN (FRAGMENT).
GN 4930577N17RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

Schröml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Schoenwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
EMBL; AK016297; BAB30183.1; -.
MGD; MGI:1914996; 4930577N17RIK.
FT NON_TER 1
SQ SEQUENCE 384 AA; 41209 MW; E9856EA599F28262 CRC64;

Query Match 2.5%; Score 10; DB 11; Length 384;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 AAAAAAAPP 149
   |||||
Db 62 AAAAAAAPP 71

RESULT 10
Q9SBM1 ID Q9SBM1 PRELIMINARY; PRT; 409 AA.
AC Q9SBM1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYDROXYPROLINE-RICH GLYCOPROTEIN DZ-HRGP PRECURSOR.
GN HRGP GENE.
OS Volvox carteri f. nagariensis.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OX Volvocaceae; Volvox.
OX NCBI_TaxID=3068;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK10;
RX MEDLINE=20044763; PubMed=10574980;
RA Ender F., Hallmann A., Amon P., Sumper M.;
RT "Response to the sexual pheromone and wounding in the green alga
RT Volvox: induction of an extracellular glycoprotein consisting almost
RT exclusively of hydroxyproline.";
RL J. Biol. Chem. 274:35023-35028(1999).
DR EMBL; AJ242540; CAB62280.1; -.
DR InterPro; IPR003882; Pistil_extensin.
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PRO1217; PRICHEXTENS.
DR PRINTS; PRO1218; PSTLEXTENSIN.
KW Signal.
FT SIGNAL 1 17
FT CHAIN 18 409 HYDROXYPROLINE-RICH GLYCOPROTEIN DZ-HRGP.
SQ SEQUENCE 409 AA; 41547 MW; CD0749C6AF02BD74 CRC64;

Query Match 2.5%; Score 10; DB 10; Length 409;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 57
   |||||
Db 129 PPPPPPPPP 138

RESULT 11
Q9FXA1 ID Q9FXA1 PRELIMINARY; PRT; 494 AA.

```

AC Q9FXA1;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE F14J22.4 PROTEIN.  
 GN F14J22.4.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,  
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,  
 RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,  
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC011807; AAG13059.1; -;  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR003592; LRR.out.  
 DR InterPro; IPR002965; P\_Rich\_extensn.  
 DR Pfam; PF00360; LRR; 8.  
 DR PRINTS; PRO0019; LEURICHRPT.  
 DR PRINTS; PRO1217; PRICHEXTENS.  
 DR SMART; SM00370; LRR; 5.  
 SQ SEQUENCE 494 AA; 54407 MW; BDC0011EEFD5C2E4 CRC64;  
 Query Match 2.5%; Score 10; DB 10; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 0.42;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 46 DCPPPPPPPPP 55  
 DB 60 DCPPPPPPPPP 69  
 RESULT 12  
 ID O61701 PRELIMINARY; PRT; 1018 AA.  
 AC O61701;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 112.9 KDA PROTEIN (FRAGMENT).  
 GN Cyk-1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Swan K.A., Severson A.F., Carter J.C., Martin P.R., Schnabel H.,  
 RA Schnabel R., Bowerman B.;  
 RT "cyk-1: a C. elegans FH gene required for a late step in embryonic  
 RT cytokinesis";  
 RL J. Cell Sci. 0:0-0(1998).  
 DR EMBL; AF062008; AAC17501.1; -;  
 DR InterPro; IPR003104; FH2.  
 DR InterPro; IPR000381; Inhibin\_beta.  
 DR Pfam; PF02181; FH2; 1.  
 DR PRINTS; PM00671; INHIBINB.  
 DR SMART; SM00498; FH2; 1.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 1018 AA; 112944 MW; 07DA53F14B000189 CRC64;  
 Query Match 2.5%; Score 10; DB 5; Length 1018;  
 Best Local Similarity 100.0%; Pred. No. 0.77;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 CPPPPPPPPPP 56  
 DB 351 CPPPPPPPPPP 360  
 RESULT 13  
 ID Q95ZV4 PRELIMINARY; PRT; 1432 AA.  
 AC Q95ZV4;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL PROTEIN F11H8.4.  
 GN F11H8.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Menezes S., Menezes S.;  
 RT "The sequence of C. elegans cosmid F11H8.";  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U40187; AAA81161.2; -;  
 SQ SEQUENCE 1432 AA; 159440 MW; 5560B3C205DBCD37 CRC64;  
 Query Match 2.5%; Score 10; DB 5; Length 1432;  
 Best Local Similarity 100.0%; Pred. No. 1;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 CPPPPPPPPPP 56  
 DB 768 CPPPPPPPPPP 777  
 RESULT 14  
 ID O02832 PRELIMINARY; PRT; 33 AA.  
 AC O02832;  
 DT 01-JUL-1997 (TREMBLrel. 04, Created)  
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE HUNTINGTIN PROTEIN (FRAGMENT).  
 GN IT15.  
 OS Gorilla gorilla (gorilla).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.  
 OX NCBI\_TaxID=9593;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96326790; PubMed=8766138;  
 RA Pecheux C., Gall A.L., Kaplan J.C., Dode C.;  
 RT "Sequence analysis of the CAG triplet repeats region in the Huntington  
 RT disease gene (IT15) in several mammalian species.";  
 RL Ann. Genet. 39:81-86(1996).

DR EMBL; S83377; AAB50771.1; -.

FT NON\_TER 1

SQ SEQUENCE 33 AA; 3393 MW; 44F558CE5636722A CRC64;

Query Match 2.2%; Score 9; DB 6; Length 33;

Best Local Similarity 100.0%; Pred. No. 0.42;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56

|||||

DB 20 PPPPPPPPP 28

RESULT 15

Q9VWS6

ID Q9VWS6 PRELIMINARY; PRT; 48 AA.

AC Q9VWS6;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)

DE CG15041 PROTEIN.

GN CG15041

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Arif J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL; AE003509; AAF48862.2; -.

DR FlyBase; FBgn0030939; CG15041.

SQ SEQUENCE 48 AA; 4976 MW; 17291770CAF8F177 CRC64;

Query Match 2.2%; Score 9; DB 5; Length 48;

Best Local Similarity 100.0%; Pred. No. 0.57;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 PPPPPPPPP 56

|||||

DB 25 PPPPPPPPP 33

Search completed: August 13, 2002, 15:15:05

Job time: 141 sec





**THIS PAGE BLANK (USPTO)**